March 2015

Stock Price Predictions

Arsene N. Tchatchoua
Worcester Polytechnic Institute

Interactive Qualifying Project

Stock Price Predictions

Department of Electrical and Computer Engineering

Student: Arsene Tchatchoua

Advisor: Humi Mayer

Student Signature: ……………………………………

Advisor Signature: ……………………………………

Date: February 27th 2015

Term: C -2015

Table of Contents

Abstract

Man desires to get rich and responsibly wealthy while working as little as possible. The playground that seems to produce this economic utopia is the stock market. In this report, we will document the development of a mathematical stock forecasting model and its performance in the prediction of 5 biotechnology stocks. After a $100,000 investment in the 5 stocks, monetary gains were made in the third week and these laid the foundations to discuss the effectiveness of our model.

Executive Summary

In the course of the model's research and development, A mathematical prediction model was built and experimented; the ARIMA/ exchange model. This model is profoundly encompassed by the linear regression technique. The Linear regression modelling technique uses the sum of a linear equation and Fourier series to depict trends in stock prices .The ARIMA/exchange model, is built upon the weighted sum of the ARIMA forecast and current market exchange indexes and when added to the regression result, we obtain our complete model.

5 biotechnology stocks were chosen on December 12th 2014 to test the ARIMA model. $100,000 was invested into these stocks in order to generate financial profits in a monthly period. Minor losses of 1.7% and 1.45% were registered in the first two prediction weeks respectively.   On the third week a profit of 2.7 % was registered and our research goal had been achieved. Nevertheless, the model's flaws were discussed and improvements were suggested for future research programs.

Introduction

The idea of stock markets originated when countries in the New World began trading with each other. While many pioneer merchants wanted to start huge businesses, this required substantial amounts of capital that no single merchant could raise alone. As a result, groups of investors pooled their savings and became business partners and co-owners with individual shares in their businesses to form joint-stock companies. Originated by the Dutch, joint-stock companies became a viable business model for many struggling businesses. In 1602, the Dutch East India Co. issued the first paper shares. This exchangeable medium allowed shareholders to conveniently buy, sell and trade their stock with other shareholders and investors. The idea made its way to the American colonies with an exchange started in Philadelphia in 1790. Undergraduates at Worcester Polytechnic Institute are required to complete two qualifying projects as degree requirements. The first project usually offered in the Junior year is the Interactive Qualifying Project (IQP) and the second project offered in the Senior year is the Major Qualifying Project (MQP).

The MQP is essentially built around the students major of study and it is concentrated on designing, conceiving and testing bright developed engineering ideas. This enables the student to apply his engineering knowledge to a real world project hence preparing him for future employment and research challenges. The IQP focuses on science and society domains outside the students major of study. This is very important in giving the student a broad perspective of social issues in the world and enables him to develop critical thinking, social skills and awareness to complete his engineering knowledge.

Why this is an Interactive Qualifying Project (IQP)?

Our Institute defines IQP and put it at best in this term "The Interactive Qualifying Project, or IQP as it is known on campus, is WPI's most distinctive academic requirement, and is unique in higher education. The IQP challenges students to address a problem that lies at the intersection of science or technology with social issues and human needs and is done under the direct guidance of one or more faculty advisors, usually I teams of 2-4 students."

An effective and organized stock market is an economic barometer that indicates a strong economy. Developed countries are all invested into stock markets because it contributes to economic growth, provides liquidity for investors (hence encouraging investors to invest in long terms with confidence) and increases the public's business literacy (by educating the public about better trading practices and advantages of ownership securities). Nevertheless, the stock market is a risky investment terrain due to its highly volatile trend behavior and this may cause developing countries and emerging markets to be sometimes timid towards stock involvement.

The objective of this IQP is predicting stock trends and behaviors over short periods of time using the science and technology available to us. By experimenting the effects of various stochastic and mathematical formulas, we strive to build a cheap, fast and accessible stock forecast model that can be understood by investors with little stock knowledge and definitely help emerging markets. Of course, in the stock exchange world, it is not recommended to treat a prediction model as a foreseeing crystal ball and close our minds about events around us. That is why in conjunction with this prediction model, this IQP also gives an insight into various economic events and seasonal trends that should be watched out when investing in the stock market.

Because our project uses technology, science and mathematics to create a tool that will help investors make informed decision before buying stocks, our project responds to the definition and objective of an IQP set by Worcester Polytechnic Institution. Therefore, we could say with confidence that this project is an IQP.

Factors affecting Stock Prices

There are hundreds of factors that will control and affect stock prices. Some of these factors are even more prominent during busy periods of the year. The figure below shows a list of a few factors

A prime example of the effect of terrorism on stock markets is illustrated in the attacks of September 11 2001.

Terrorism Attack of September 11, 2001

Market reaction

"Anticipating market chaos, panic selling and a disastrous loss of value in the wake of the attacks, the NYSE and the Nasdaq remained closed until September 17, the longest shutdown since 1933. Moreover, many trading, brokerage and other financial firms had offices in the World Trade Center and were unable to function in the wake of the tragic loss of life and collapse of both towers.

On the first day of NYSE trading after 9/11, the market fell 684 points, a 7.1% decline, setting a record for the biggest loss in exchange history for one trading day. At the close of trading that Friday, ending a week that saw the biggest losses in NYSE history, the Dow Jones was down almost 1,370 points, representing a loss of over 14%. The Standard and Poor's (S&P) lost 11.6%. An estimated $1.4 trillion in value was lost in those five days of trading. Major stock sell-off hit the airline and insurance sectors as anticipated when trading resumed. Hardest hit were American Airlines and United Airlines, carriers whose planes were hijacked for the terrorist attacks.

The Financial Aftermath

American Airlines (NYSE:AMR) stock dropped from a $29.70 per share close of September 11 to $18.00 per share close on September 17, a 39% decline.

United Airlines (NYSE:UAL) stock dropped from $30.82 per share close to $17.50 per share on the close on September 17, a 42% decline. Similar steep declines hit the travel, tourism, hospitality, entertainment and financial services sectors, as a wave of temporary fear and uncertainty swept through the nation. Among the financial services giants with the steepest drops in share prices were Merrill Lynch which lost 11.5%, and Morgan Stanley which lost 13%. Insurance firms reportedly eventually paid out some $40.2 billion in 9/11 related claims. Among the biggest losers was Warren Buffet's Berkshire Hathaway. Most insurance firms subsequently dropped terrorist coverage".

Research

My research was subdivided into two parts. The first part depicts the performance of 10 automobile stocks used to analysis the linear regression model.

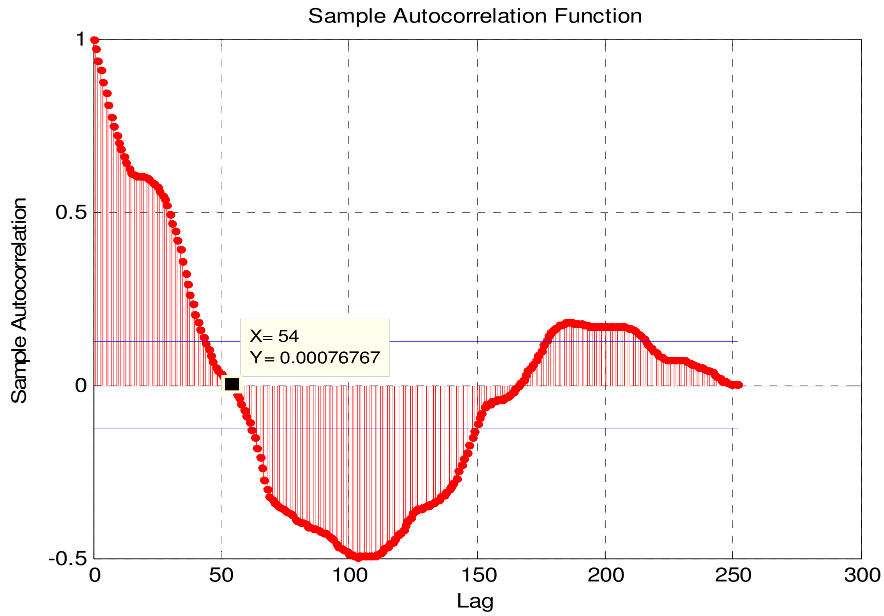
The Linear Regression Model

The linear regression technique was used to follow the prices trend over a year and propose future prices. 10 automobile stocks were chosen to test the linear regression technique. In order to bolster the regression line across each automobile stock, autocorrelation was applied on 286 sample stock prices of each Automobile stock (September 12th 2013 to September 12th 2014). The positive correlated periods obtained were used as our regression line inputs. A positive correlation meant that the sample values tended to follow a common tandem; as one sample value increases the next value increases, when one sample decreases the next sample value also decreases.

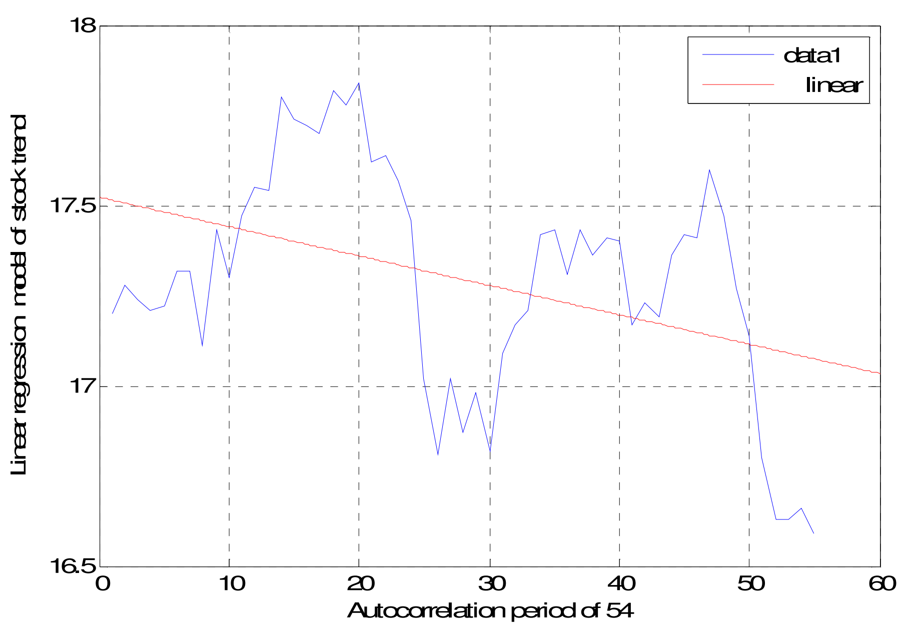
The 10 automobile stocks chosen are shown below

Symbol	Name	Last Trade	Type	Exchange
F	Ford Motor Co.	16.33	Stock	NYSE Currency in USD
NSANY	Nissan Motor Co. Ltd	19.73	Stock	Currency in USD
TM	Toyota Motor Corporation (TM)	118.65	Stock	NYSE Currency in USD

HMC	Honda Motor Co Ltd	34.42	Stock	NYSE Currency in USD
GM	General Motors Company	33.17	Stock	NYSE Currency in USD
THO	Thor industries Inc.	52.92	Stock	NYSE Currency in USD
WGO	Winnebago Industries, Inc	23.51	Stock	NYSE Currency in USD
TTM	Tata Motors Limited	44.68	Stock	NYSE Currency in USD
RNSDF	Renault Soci	74.98	Stock	Other OTC Currency in USD
HYMTF	Hyundai Motor Company	56.35	Stock	Other OTC Currency in USD

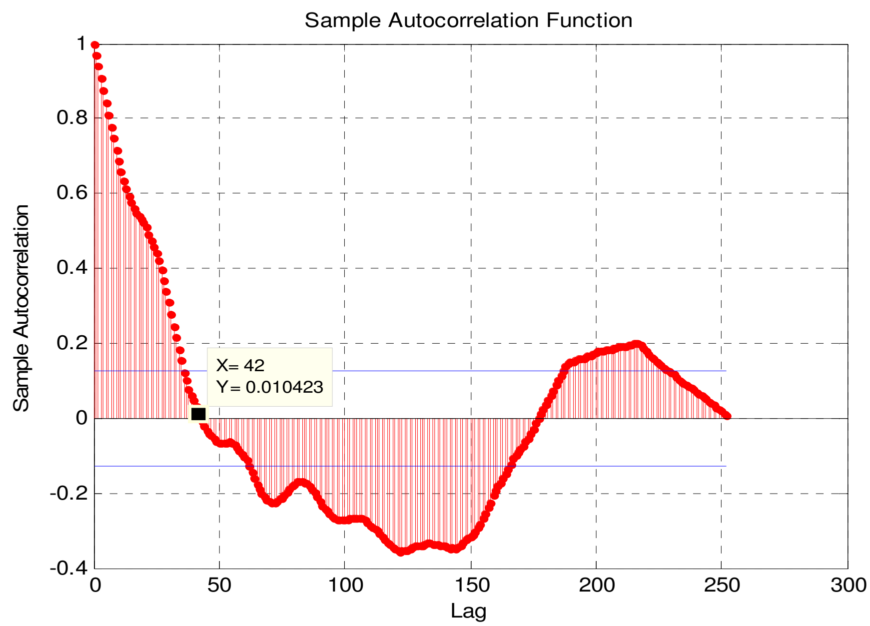


Autocorrelation for Ford motor

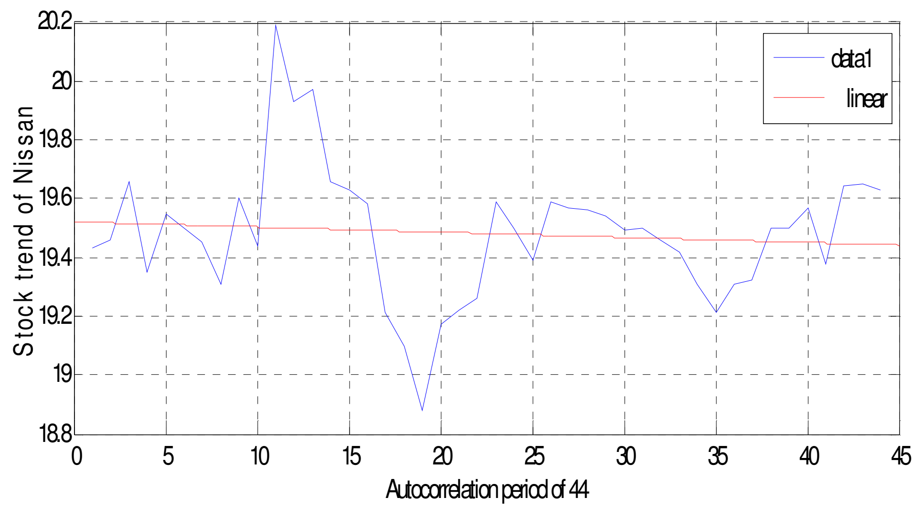


Linear regression of Ford

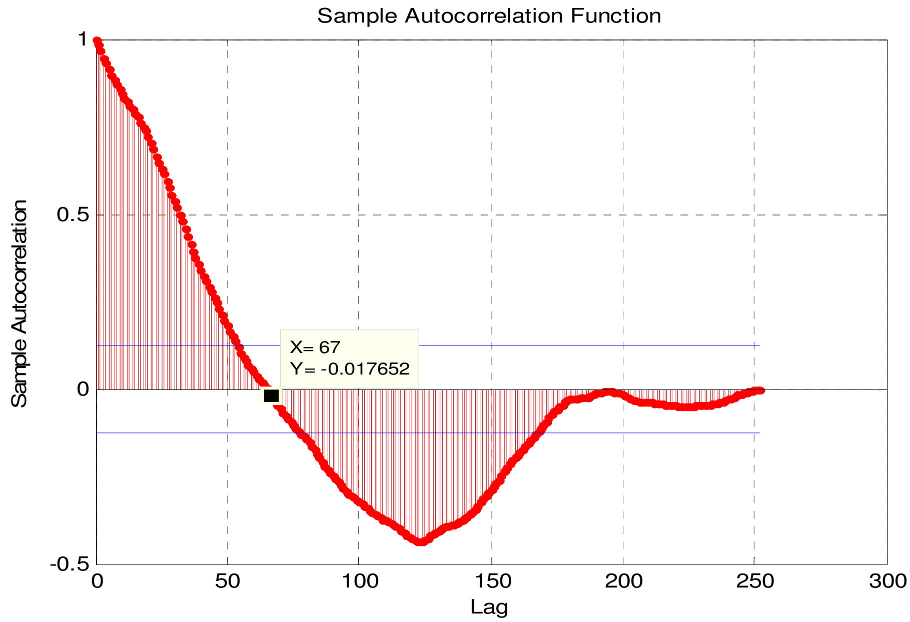
y = -0.0081522) t + 17.523



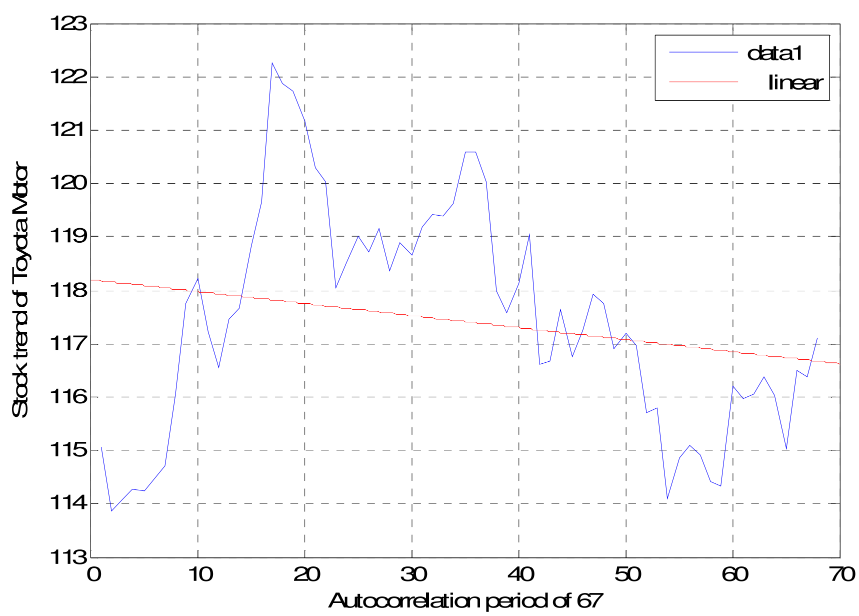
Autocorrelation for Nissan Motor



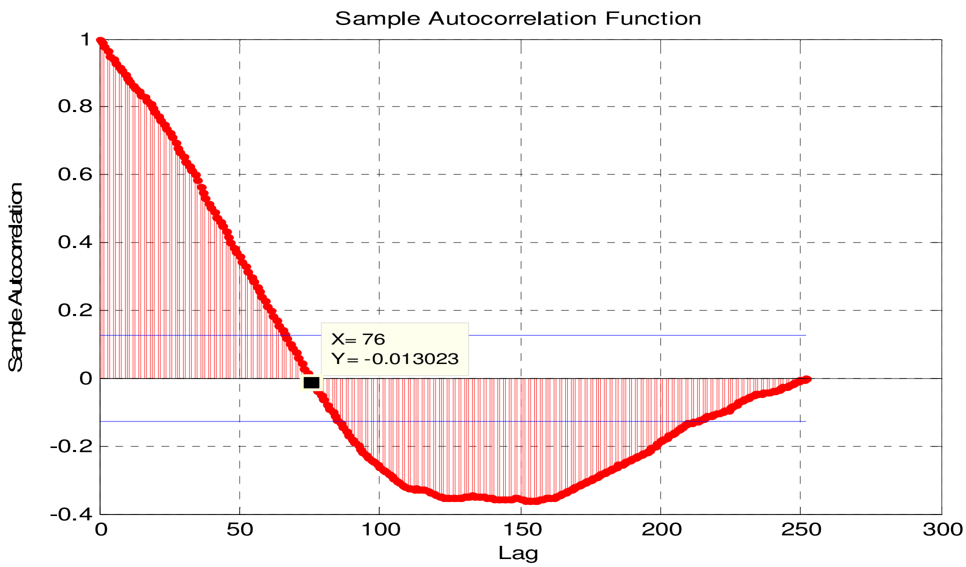
y = -0.0017604*t + 19.521



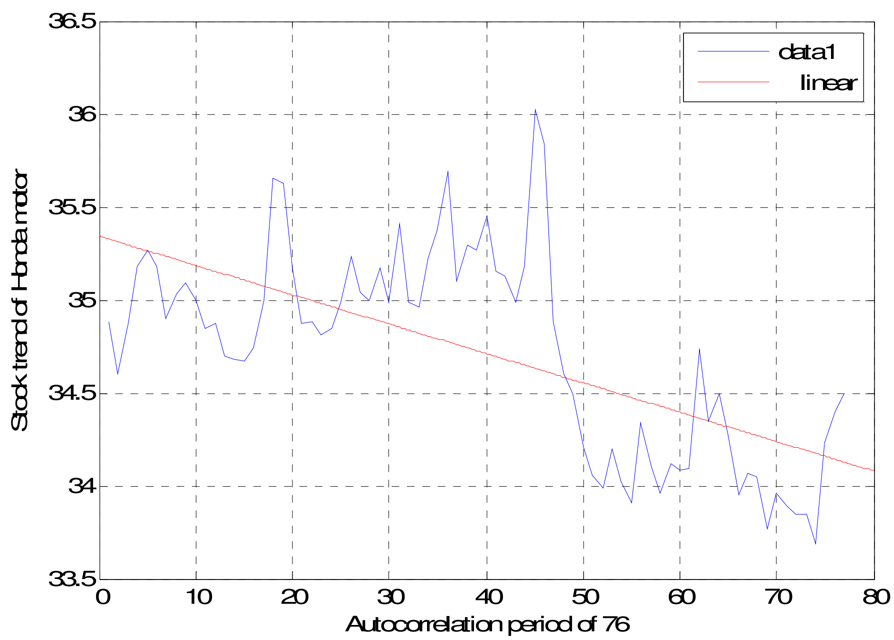
Autocorrelation for Toyota Motor



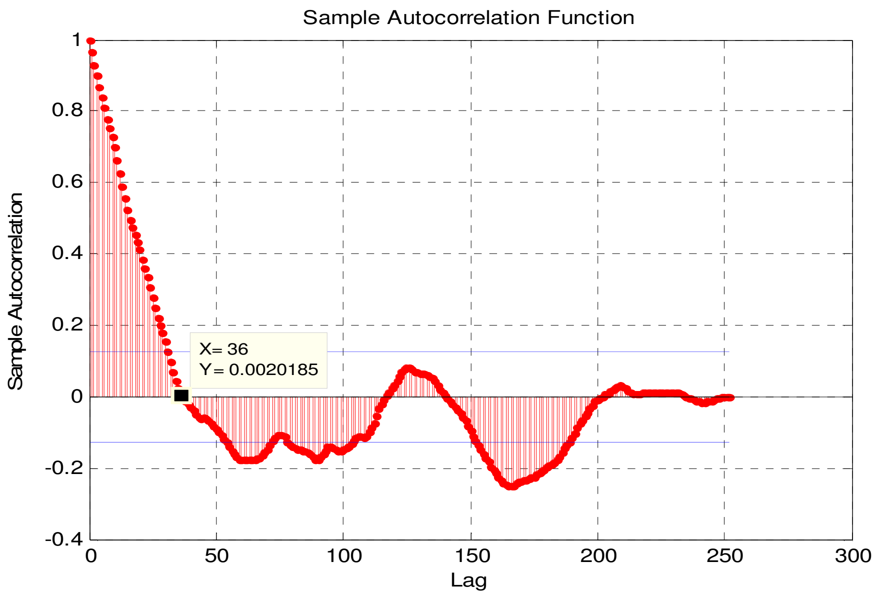
Y = -0.022538*(t) + 118.2.



Autocorrelation for Honda Motor

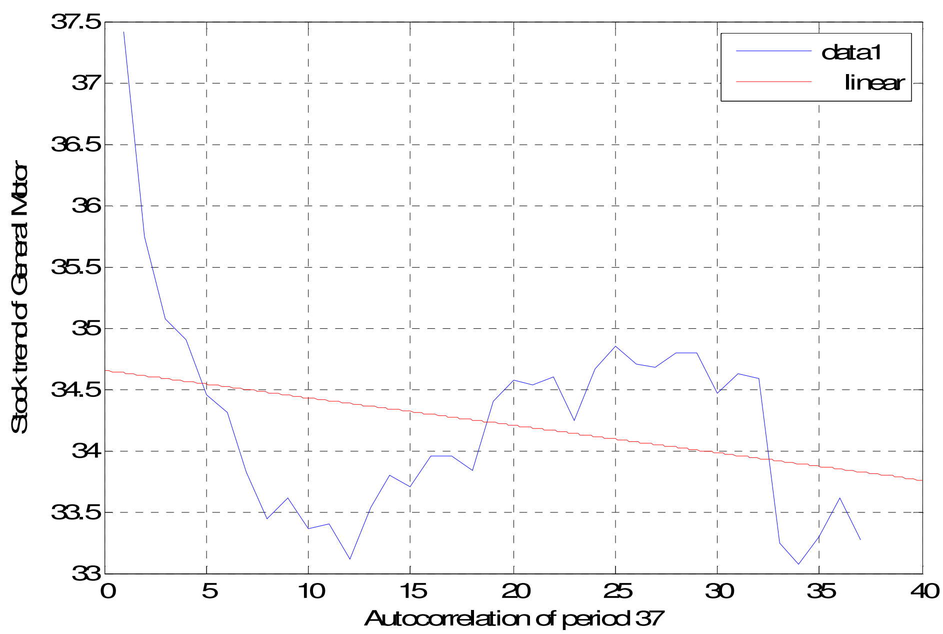


$y = (-0.015796)*t+ 35.343;$

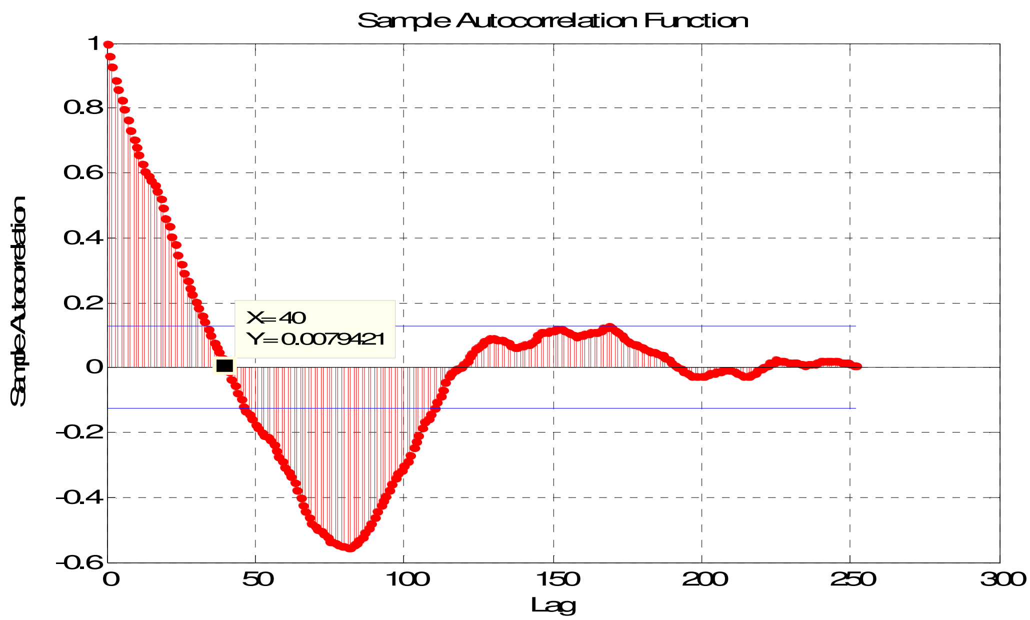


Autocorrelation for General Motor

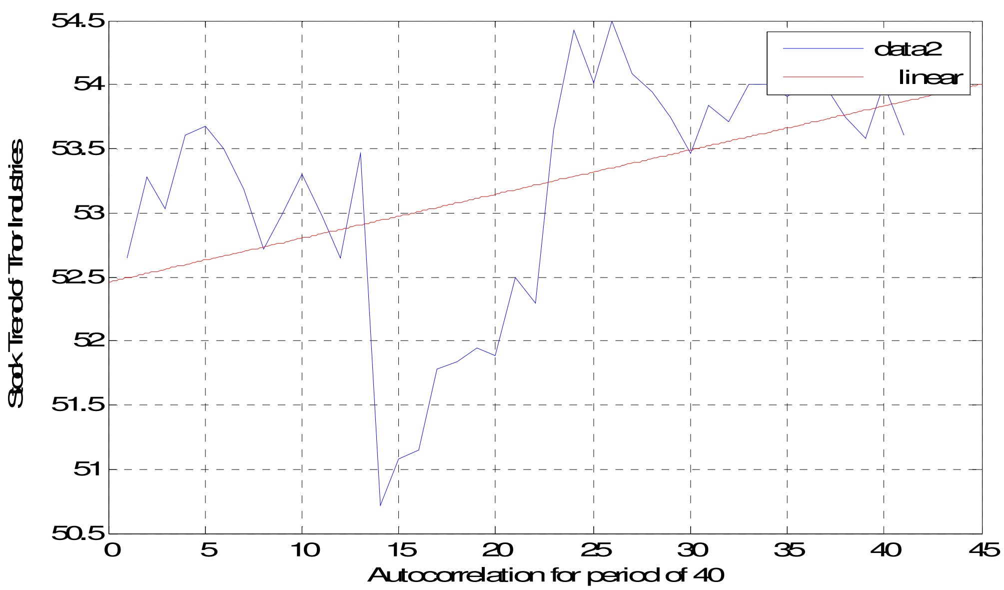
y = (-0.022397)*t + 34.653



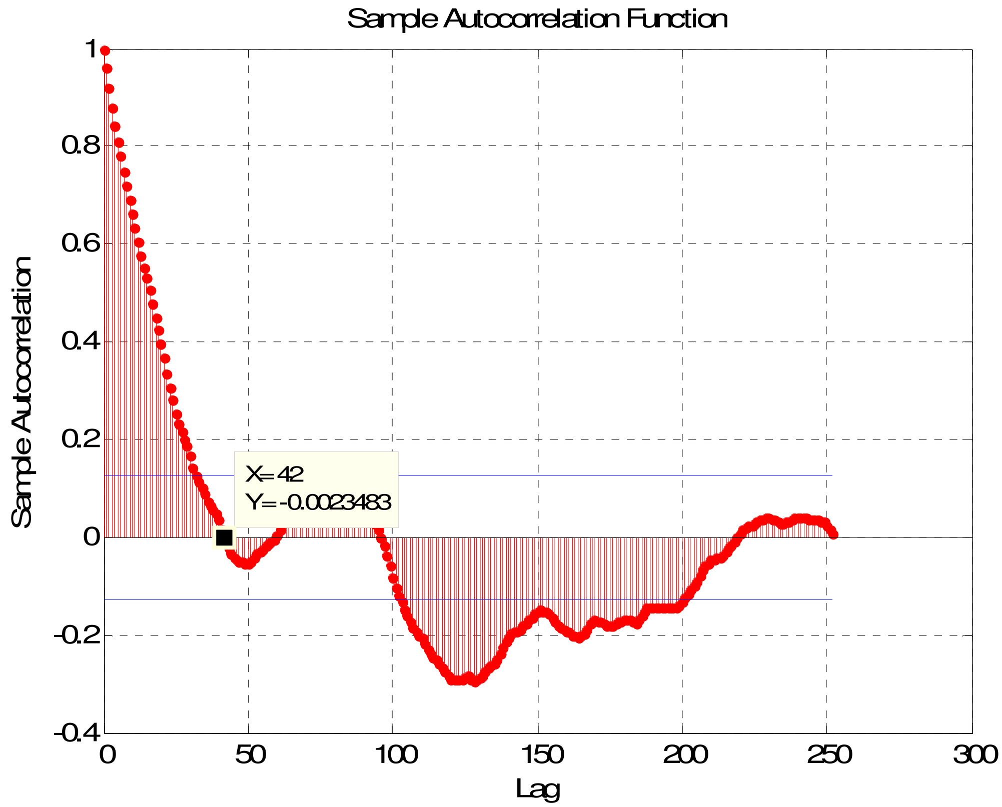
General motor



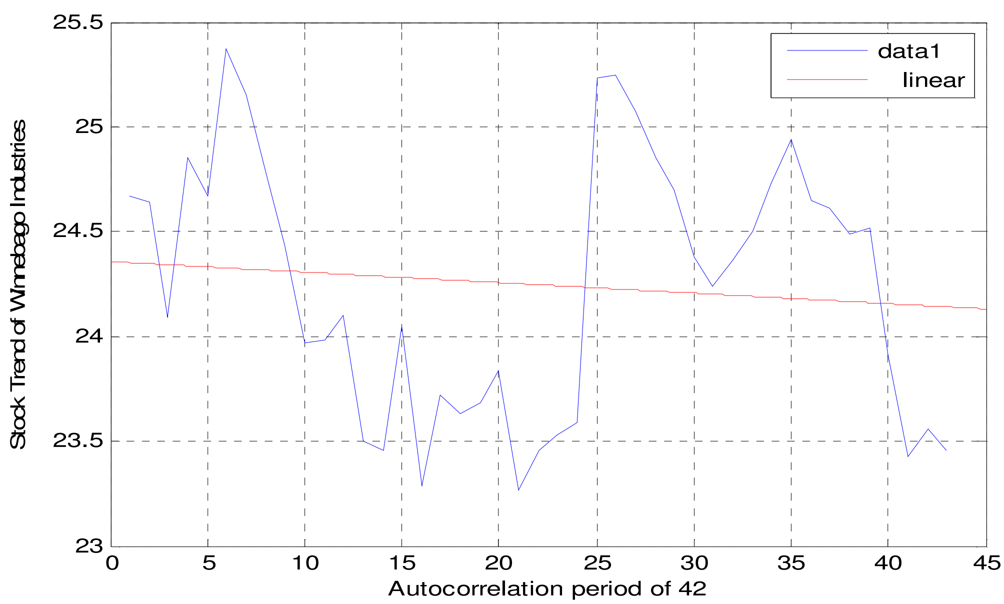
Autocorrelation for Thor Industries



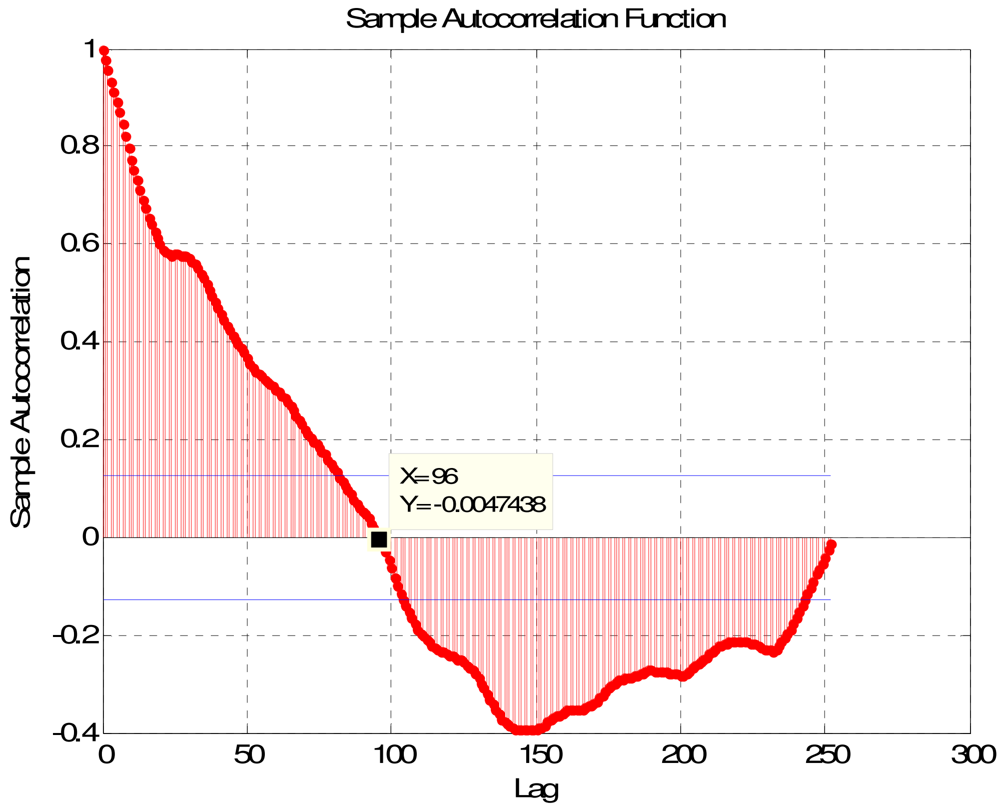
Y = (0.034392)*t + 52.458;



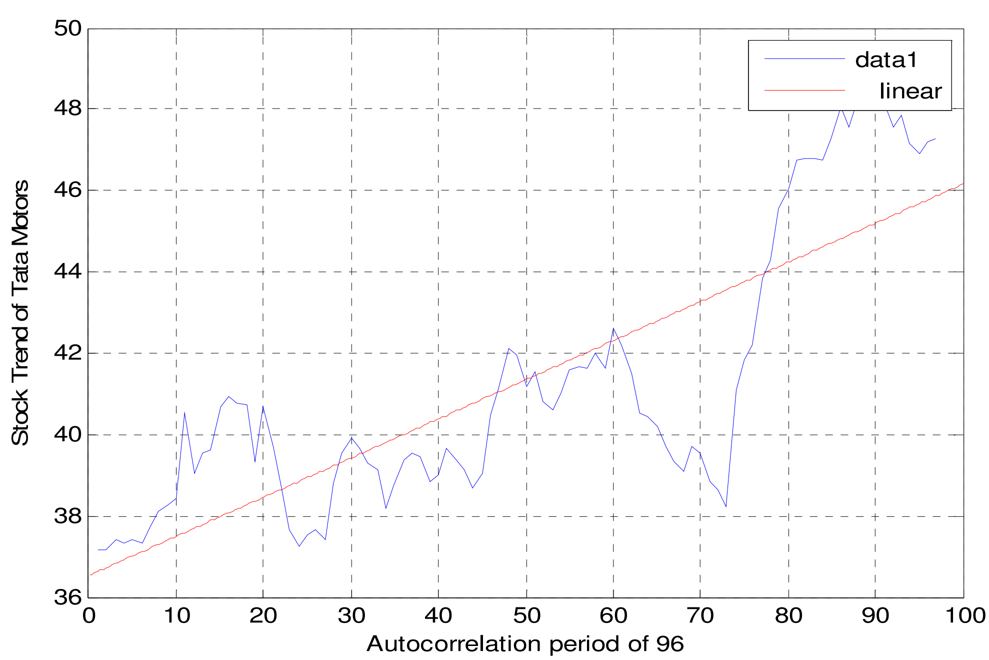
Autocorrelation for Winnebago Industries



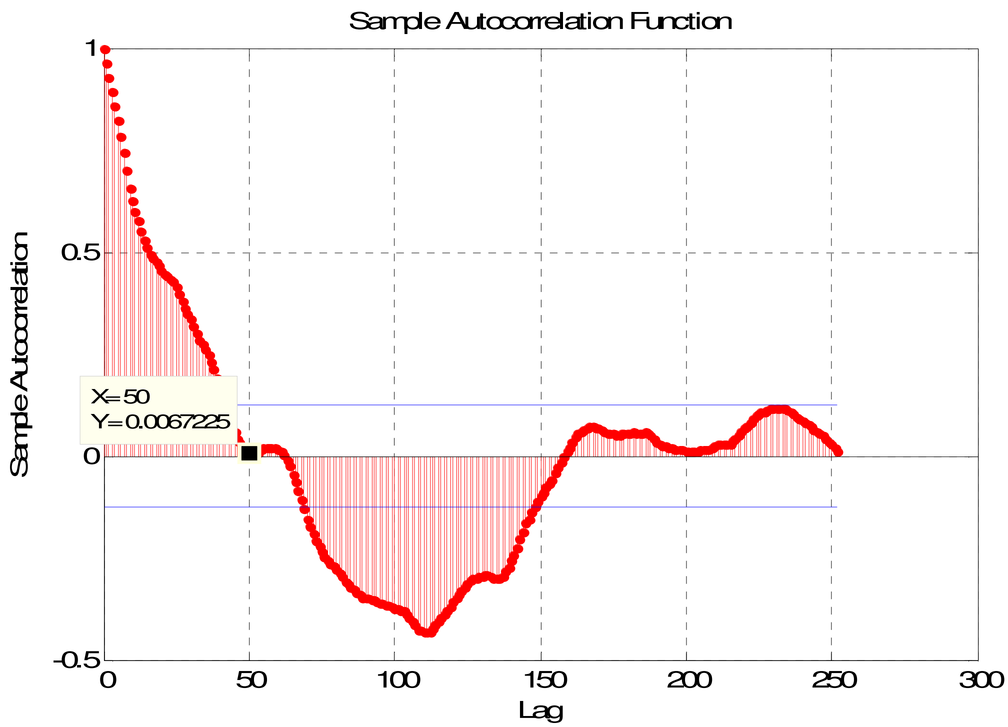
Y = (-0.0050121)*t + 24.357;



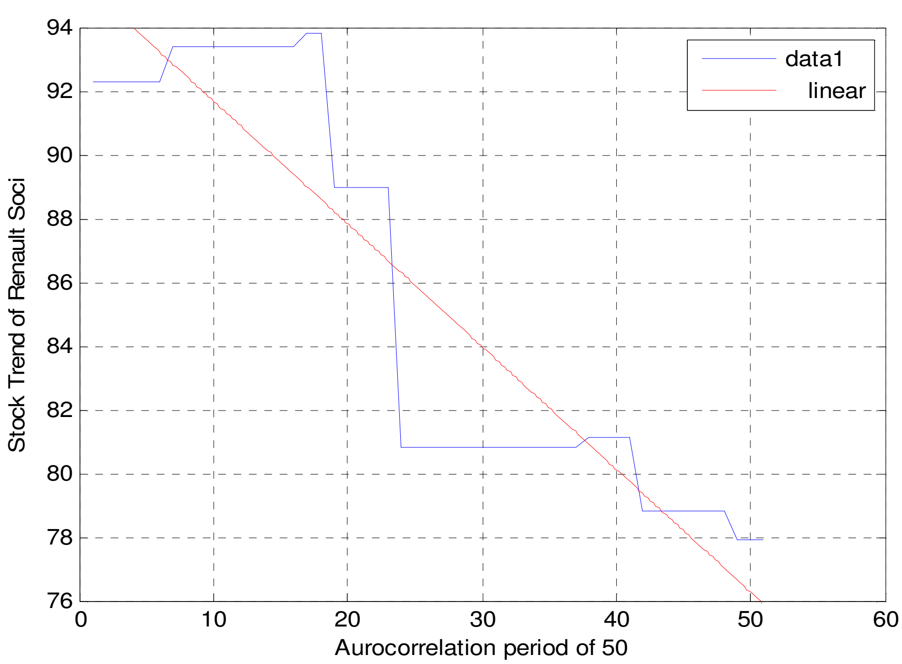
Autocorrelation for Tata Motors



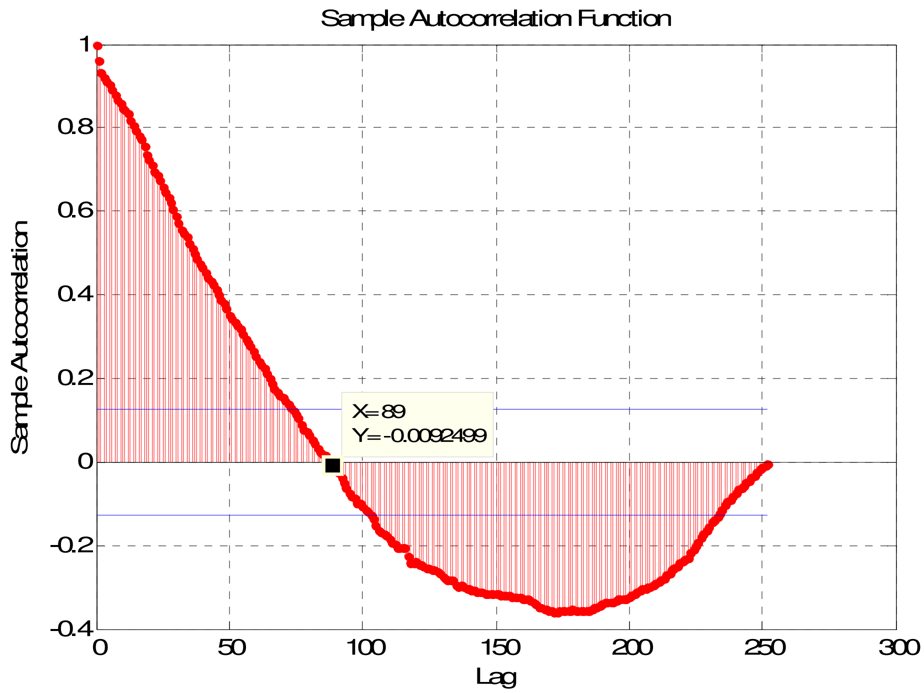
Y = (0.096273)*t + 36.534;



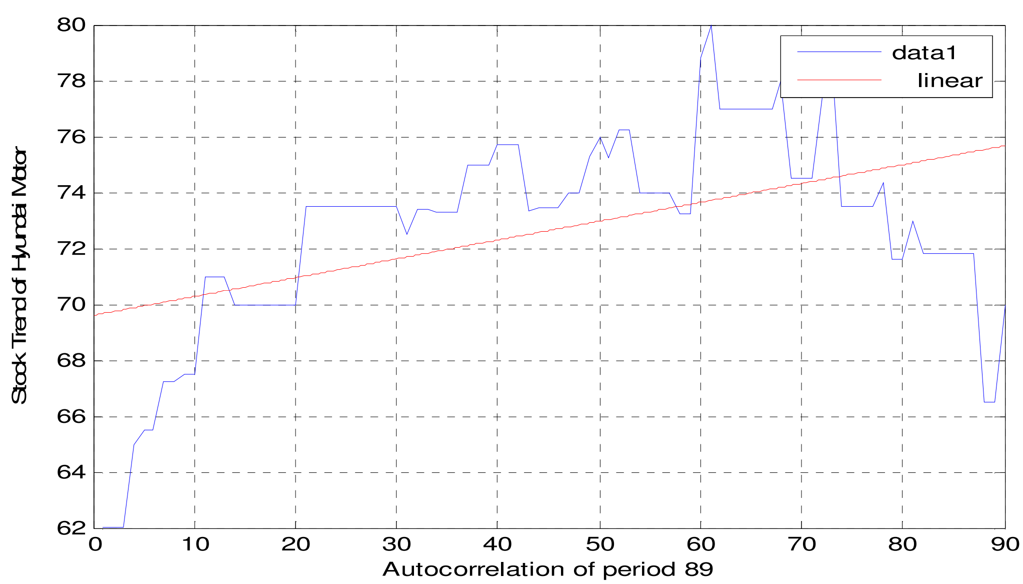
Autocorrelation for Renault Soci



Y = (-0.38539)*t + 95.546



Autocorrelation for Hyundai Motor



Y = (0.067357)*t + 69.604

The regression technique draws a line through a graph that best fits and describes the moving behavior of the curve. The regression equation is

$$\boxed{Y=P1x + P2}$$

Where **P1** and **P2** are called the regression coefficients and **a** is the slope of the regression line while b is its y intercept. The Mat lab basic fitting function was used to adequately draw the regression line through each curve and the respective future periods were substituted in x to get each predicted price. Lets take a look at Tata motors stock trend.

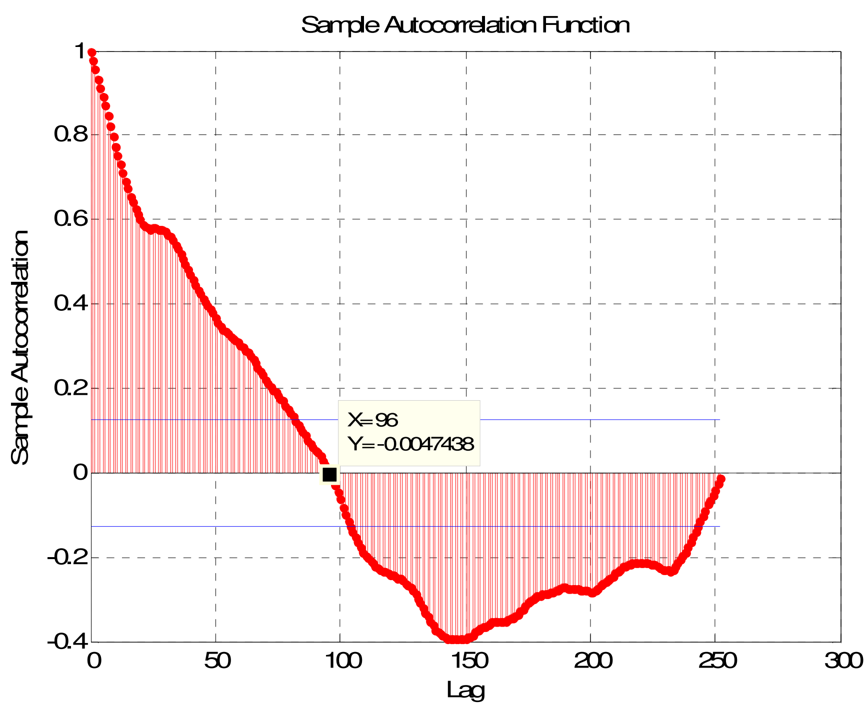


Figure 1a: Autocorrelation for Tata Motor

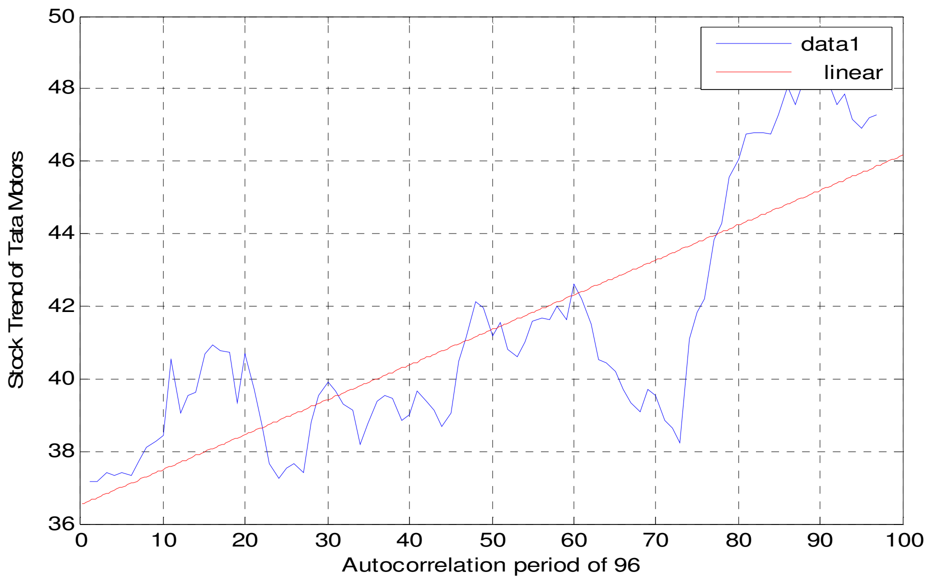


Figure 1b: Linear Regression for Tata Motor

Autocorrelation was applied to Tata motor stock prices between September 12th 2013 and September 12th 2014. As shown in the figure 1a, a positive correlation period of 96 was obtained and this period was used to draw the regression line in figure 1b above. We can see the line best fits and describes the curve because the sample values all move in a common tandem.

This is one advantage of the regression line technique. It takes into consideration the best sample values for each stock that will effectively produce the best results. The regression line will move along the curve and produce linear coefficients which are m and **b** in

$$\boxed{\underline{\mathbf{Y=mX+b}}}$$

Where **m** is the gradient of the line and **b** is the y intercept of the line. The gradient of the line has to cover the entire stock trend to adequately model it. Additionally, the y intercept of the line has to be as close as possible to that of the curve to minimize any offsets in the y intercept coefficients. These coefficients will then be used to predict values for future sample values.

The problem with the regression line is the fluctuating nature of stock prices. Stock trends to do not follow a linear path and the error between our regression line and the actual stock trend tend to affect our regression coefficients. Let's consider the regression models for Ford industries and Nissan motor respectively:

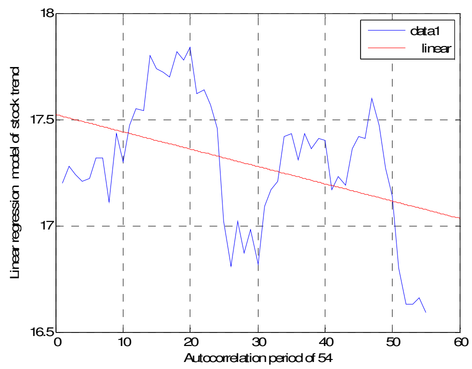


Figure 2a: Ford industries regression line

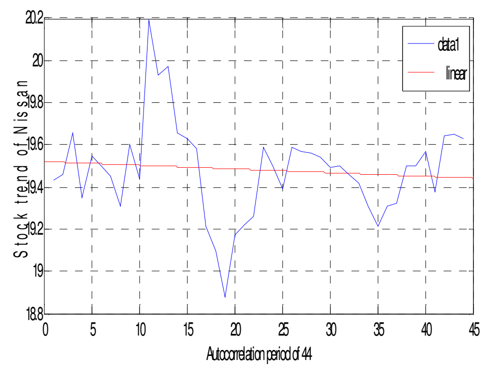


Figure 2b: Nissan motor regression line

We can see in the figures above that each regression line fits and follows its respective stock trend. Nevertheless, there is a significant difference between the two graphs. In Figure 2a, the line tends to fit the average movement of the curve as the curve fluctuates about the line. The model is good till it reaches period 50 where the stock trend undergoes a sharp dip. Since the regression line considers the average movement of the stock trend, the sharp dip at the end of the curve is not taken into consideration. Consequently, the gradient of the regression line is

less than it has to be to effectively cover the stock trend. Additionally, in figure 2a, the line's y-intercept is higher than that of the curve and this will account for a much higher **b** coefficient than required.

In Figure 2b, the curve tends to have many fluctuations but the slope of the regression line effectively spans the entire stock trend and the line's y intercept is very close to that of the curve. Hence, the coefficients of the line for figure 2b are a more descriptive measure of the actual stock trend.

The Fourier Series

Baron Jean Baptiste Joseph Fourier (1768 − 1830) introduced the idea that any periodic function can be represented by a series of sines and cosines which are harmonically related.

To consider this idea in more detail, we should introduce some definitions and common terms.

All periodic functions can be represented by sets of sine and cosine functions and sometimes exponential functions. This periodic functions are operated by the Fourier Transform which is a bilateral case of the laplace transform with s=2πiε..ie

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)\ e^{-2\pi i x \xi}\, dx$$
,

This set of sines and cosines consist of the Fourier series. The Fourier series is a set of sums of cosine and sine functions where the frequency of each successive term is a positive integer multiple of the fundamental frequency of the first term.

$f(x)$ = 2cos2t + 3 cos4t + 6cos6t + 8cos8t….

where F_o=fundamental frequency and we can see =2

,F_1=2(2) =4 , F_2=3(2)=6….

The general form of the Fourier series is

$$f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L}\right) \quad \text{where n is any positive interger}$$

Fourier analysis was applied to the error between each stock trend and its respective regression line. The Fourier series breaks down a time signal into a Taylor sum of cosines and sinusoids. This feature is very critical to our model because only the main signal in the error will be expressed in the Fourier domain. Noise has a high frequency so it will be filtered out of the error because the cosines and sinusoids require a signal to have a significant time period. In general, the Fourier series is defined as

$$f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L}\right)$$

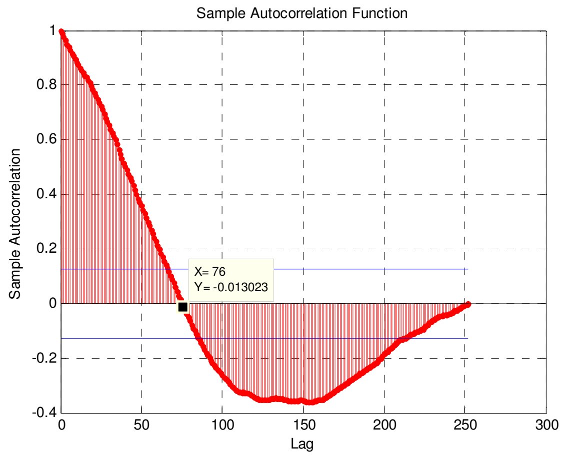
Where ao is the constant term coefficient, an is the cosine coefficient, bn is the sinusoid coefficient, n is the index number and L is period/2

We notice in the definition that the Fourier series definition considers values of n for n = [1: ∞] because stock trends are real and casual signals, hence there can't be index values for n<0

The Fourier coefficient depends on three factors, the movement of the stock prices, and the magnitude of the error between the stock trend and its regression line and the number sample values used.

The movement of the stock prices determines the stock trend's shape and how the Fourier equation will break down the trend into sinusoids. The stocks movements are caused by human factors that cannot be determined. Nevertheless, the magnitude of the error can be determined. An increase in the error's magnitude will cause the Fourier coefficients to be large and less accurate with a large noise influence on the resulting sinusoids. A decrease in the error's magnitude will produce Fourier coefficients with smaller values. We want our Fourier coefficients to be as small as possible.

The number sample values used in the Fourier analysis is very important. We used Fourier 1, 2 and 3 for our model. Fourier 1 does not require a lot of values because it defines only three coefficients but Fourier 2 and 3 require more sample values because more coefficients are to be found. Let's consider the Fourier 3 model of the error for Honda motor and General motor. Honda motor and General motor have respective average closed prices of $ 34.42 and $ 33.17.



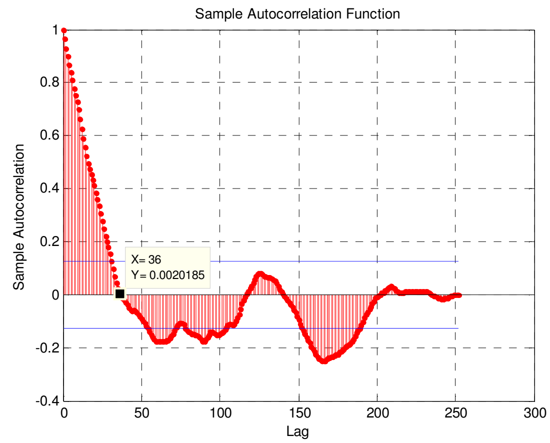
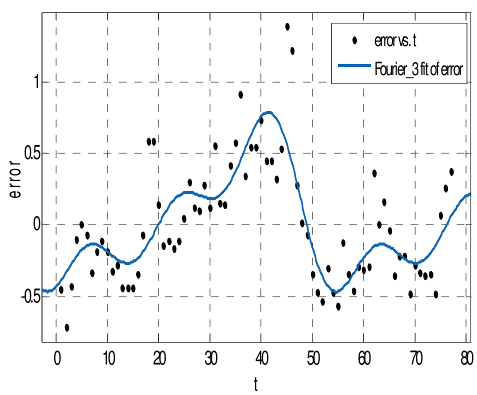


Figure 3a: Positive correlation of 76 for Honda

Figure 3b: Positive correlation of 36 for General Motor



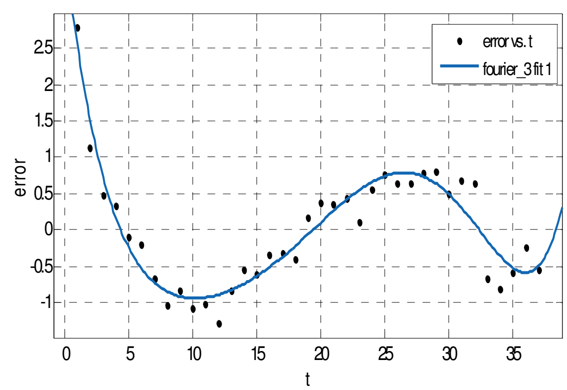


Figure 4a: Fourier 3 of error for Honda

Figure 4b: Fourier 3 of error for General Motor

In figure 3a, we can see that the correlation period of Honda Motor is 76 while in figure 3b that of General Motor is only 36. In effect let's consider the respective Fourier 3 coefficients produced for each stock:

Fourier 3 Coefficients(with 95% confidence bounds):	Honda Motors	General Motors
w	0.1118	-0.004214
a_0	0.07051	4.83e+08
a_1	-0.2716	-7.224e+08
b_1	-0.333	5.655e+07
a_2	-0.1847	2.864e+08
b_2	0.05266	-4.512e+07
a_3	-0.04851	-4.705e+07
b_3	0.1896	1.123e+07

Table 1: Coefficients of Fourier 3 for Honda and General Motors

We can see from the table above that even though these two stocks have very similar close prices (34.42 and 33.17), they produce very different Fourier coefficients. The positive correlated period of Honda motor is 76 so it had enough sample values to effectively model Fourier 3 and produce coefficients with small values. But General motors had a positive correlated period of 36 hence most of the noise in Fourier 3 was not filtered, and the coefficients produced had abnormally large values. In fact, this result shows us that Fourier 3 is not a good model to filter out the noise from the error in General motor trend because the stock trend does not have enough input values.

Exceptions do occur like the Fourier 2 and 3 of Toyota motors. Fourier 2 tends to produce abnormally large values while Fourier 3 produces normal values. Toyota has a positive correlation period of 67.

Toyota Motors	**Fourier2**	**Fourier3**
w	-0.0009423	0.07664
a_0	1.681e+07	-0.3595
a_1	-2.241e+07	-1.656
b_1	2.933e+05	1.361
a_2	5.609e+06	-1.021
b_2	-1.47e+05	-0.6985
a_3	-	-1.04
b_3	-	-0.4439

Table 2: Fourier2 and Fourier3 comparison for Toyota Motors

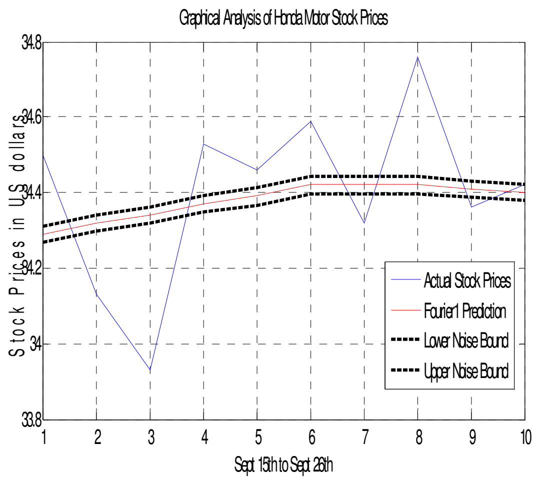
The last step in our linear regression graphical model is to sum up the linear regression prediction to the Fourier prediction. By summing these two entities, we add a noise filtered error bound to the linear regression prediction to make the final prediction more robust. It will be smart to add a noise bound to our linear regression prediction but sometimes the linear prediction is already fine on its own and the addition of a noise bound will decrease the prediction's accuracy. Let's consider the final predictions for the Fourier model of Honda motors and General motors:

Time (September)	**15th**	**16th**	**17th**	**18th**	**19th**	**20th**	**21th**	**22th**	**23th**	**24th**	**25th**	**26th**
Linear regression	34.0793	34.0635	34.0477	34.0319	34.0161	34.0003	33.9845	33.9687	33.9530	33.9372	33.9214	33.905
Fourier_1	0.2104	0.2539	0.2953	0.3338	0.3691	0.4006	0.4280	0.4508	0.4689	0.4818	0.4896	0.4920
New Stock Prediction	34.2897	34.3175	34.3430	34.3658	34.3852	34.4010	34.4125	34.4196	34.4218	34.4190	34.4109	34.3997
Actual Stock Prices	34.50	34.13	33.93	34.53	34.46	_	_	34.59	34.32	34.76	34.36	34.42

<u>Table 3: New Stock values of Honda motors for Fourier 1</u>

Time (September)	**15th**	**16th**	**17th**	**18th**	**19th**	**20th**	**21th**	**22th**	**23th**	**24th**	**25th**	**26th**
Linear regression	33.7571	33.7347	33.7123	33.6899	33.6675	33.6451	33.6227	33.6003	33.5779	33.5555	33.5331	33.5108
Fourier_1	-0.9805	-0.9197	-0.8097	-0.6563	-0.4679	-0.2546	0.0277	0.2005	0.4179	0.6126	0.7743	0.8943
New Stock Prediction	32.776	32.815	32.903	33.034	33.199	33.391	33.595	33.801	33.999	34.168	34.307	34.405
Actual Stock Prices	33.63	33.71	33.85	34.03	33.94			33.44	33.22	33.65	32.87	33.17

<u>Table 4: New Stock values of General motors for Fourier 1</u>



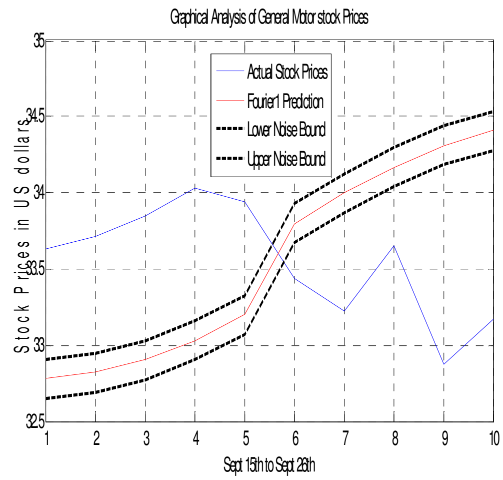


Figure 5a: Graphical prediction of Honda

Figure 5b: Graphical prediction of General M

In table 3, the linear regression values of Honda motors are far off from the actual stock prices but when the Fourier 1 values are added, the new prediction values are very close to the actual stock prices. The opposite is seen for General Motors. We can see in table 4 that the linear prediction values are already close to the actual stock prices but when the Fourier 1 values are added, the new prediction values become far off from the actual stock prices. These observations can be shown in Figure 5a and 5b respectively.

The linear regression/ Fourier graphical model depends on the performance of the linear regression model and the Fourier model used. The linear regression line needs to have a gradient that spans the stock trend and a y intercept very close to the curve's intercept in order to give more accurate prediction values. Even though we may have an accurate linear regression line, sharp dips and rises at the edge of stock trends; due to the unpredictable behavior of stock prices, may still account for significant errors. The positive correlation period of the stock trend is the critical factor that determines which Fourier model is best used. All the automobile stocks had enough sample values to model Fourier 1, 2 and 3 but some stocks like Hyundai motor and General motors did not have enough values to model Fourier 3 so Fourier 2 was the best model for them. Fourier 2 produces the best results for most stocks because it requires input values between Fourier 1 and Fourier 3 and most correlation periods fall within that range.

Adding the Fourier model to the linear regression will not always produce better values.

If a young investor was to try this model, the stocks he will use will have to have

1) A large correlation period
2) A low to average amount of stock fluctuations over the given sample period

The ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) is a statistical model that uses time series data to analysis and forecast future trends. This model seeks to predict future movements along seemingly random walks taken by stocks by examining the differences between values in the series instead of using the actual data values. Lags of the difference series are called autoregressive and lags within forecasted data are referred to as moving average. For a time series to be adequately modelled by ARIMA, it needs to be stationary. A stationary time variable is one whose statistical properties are all constant over time. A stationary series has no trend, its variations around its mean have constant amplitude, and it wiggles in a consistent fashion. Practically, real time series and events are not stationary. To correct this discrepancy, ARIMA enables the addition of a non-seasonal parameter to make time series stationary. ARIMA is depicted by the formula ARIMA(p,q,d), where p is the number of autoregressive terms, d is the number of non-seasonal differences needed for stationarity and q is the number of lagged forecast errors in the prediction equation. The ARIMA(3,1,3) model was used in our stock forecast model.

During the course of our research, it was observed that ARIMA was not an efficient forecasting tool when used as an unaccompanied model. The figures below show the various forecasting results of ARIMA(3,1,3) used to predict the prices of Nissan and Toyota motor

;

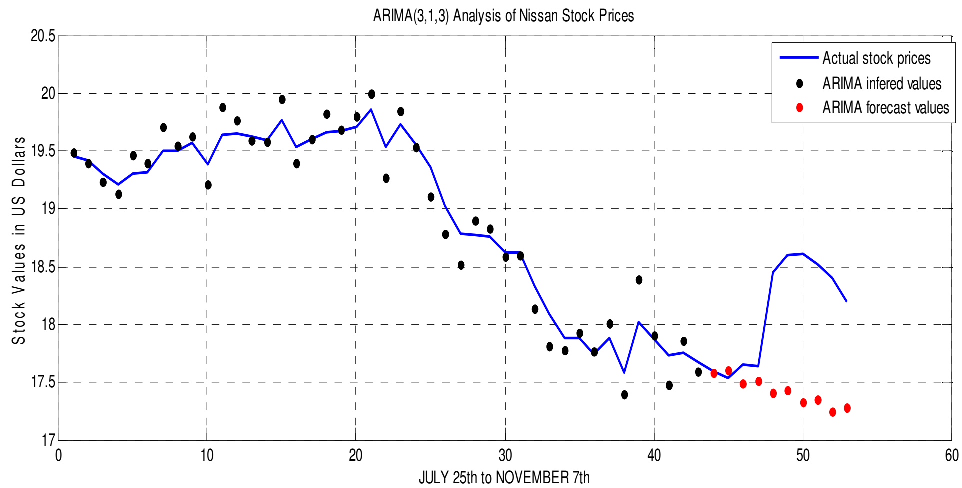


Figure 6a: ARIMA (3,1,3) analysis of Nissan stocks for 10 days

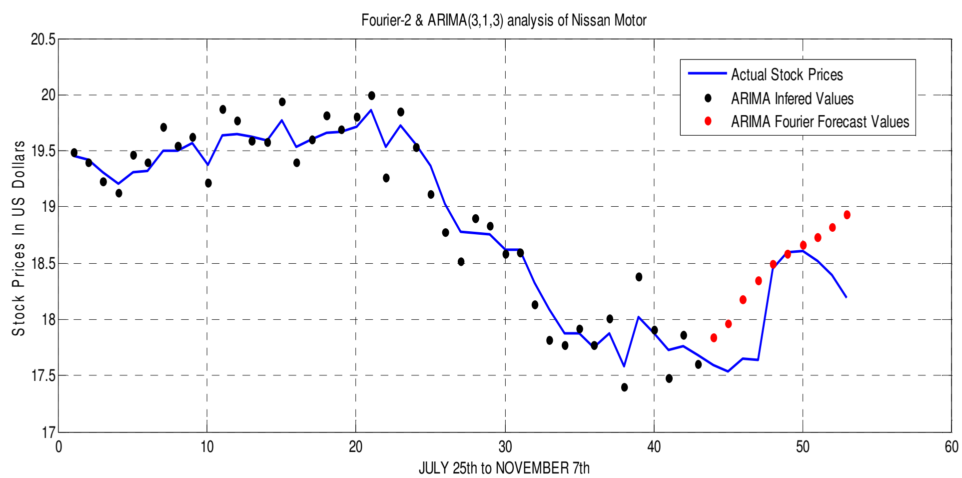


Figure 6b: ARIMA (3,1,3) & Fourier-2 model analysis of Nissan stocks for 10 days

As seen in both figures above, the ARIMA model could efficiently estimate and display the past stock values but differed in their forecasting. In figure6a, the ARIMA prediction follows the course of the last past data value which leads to a very linear forecast graph. In figure 6b, where fourier analysis was added to ARIMA, the forecasting model is less linear and does not follow the course of the last past data point. A similar result can be observed from the analysis of Toyota Motor as shown below

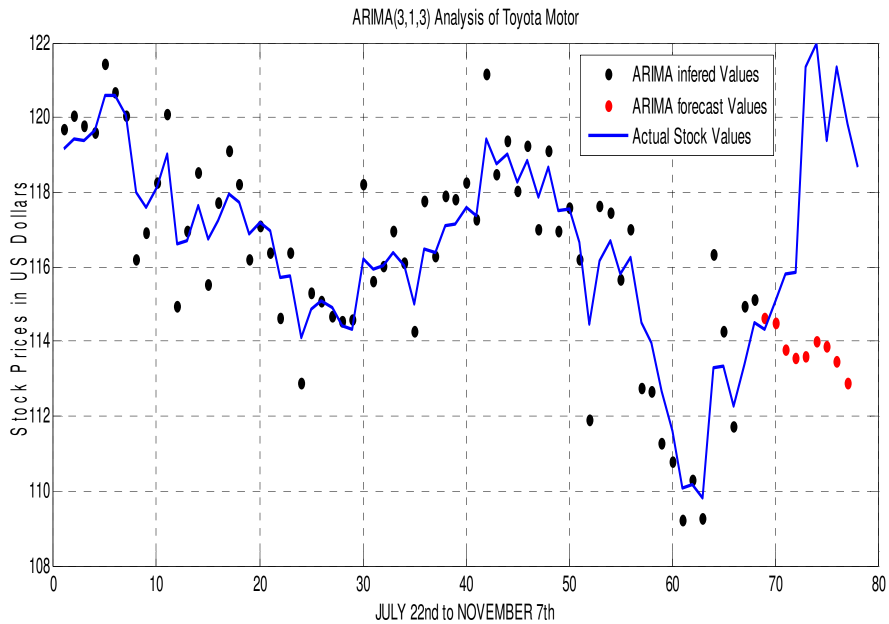


Figure 7a: ARIMA (3,1,3) analysis of Toyota stocks for 10 days

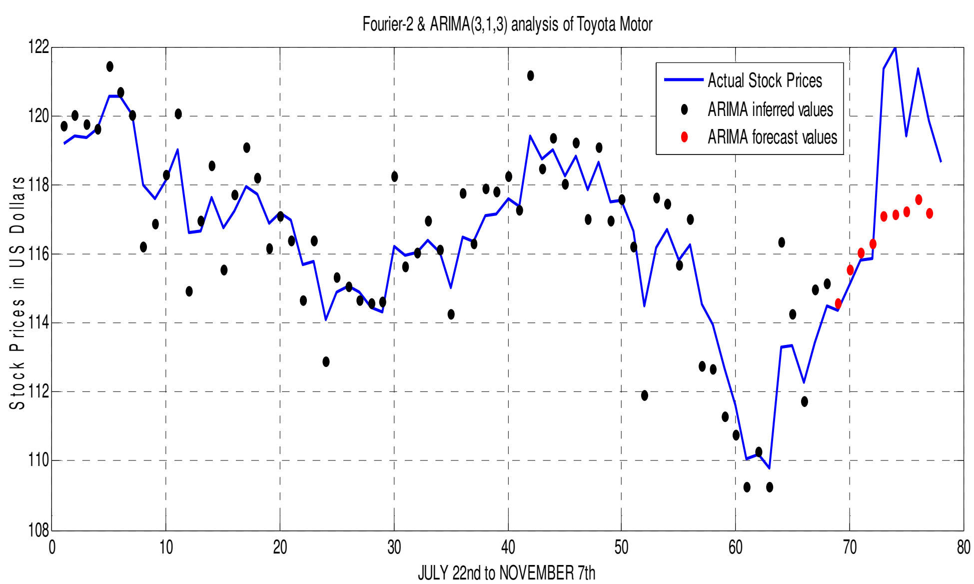


Figure 7b: ARIMA (3,1,3) & Fourier-2 model analysis of Toyota stocks for 10 days

The addition of the Fourier analysis tends to make the model more sinusoidal and decreases the linear trend of ARIMA. From these observations, we decided to always sum the fourier analysis of a a time series to its ARIMA prediction to produce the complete stock forecasting model.

Stock Forecasting Model.= ARIMA(3,1,3) prediction + FOURIER Analysis

Three factors were used to choose the five biotechnology stocks used for the ARIMA prediction.

The Price Volatility Of The Stock

The Volatility of each stock was measured by dividing each stock's standard deviation by its respective average mean value.

$$\textbf{Volatility} = (\sigma/\mu)$$

Where σ = Standard Deviation of the stocks historical data, μ = mean of the stock's historical data. Our main forecasting model consisted of superimposing The ARIMA(3,1,3) model with a NYSE/NASDAQ market model. The ARIMA(3,1,3) and Fourier 2 analysis were used to forecast each stock prices and linear regression and Fourier 2 analysis were applied unto each stock's market exchange index. The two results were then summed up in the equation below;

$$\textbf{Model} = [(1-\mu)*Ms + (\mu)*Me]$$

Where Ms = ARIMA Stock Forecast model, Me = Market exchange model, μ = Summing constant

Through repeated calculus differentiation of the model using MATLAB, the summing constant, μ was determined to be 0.4

Summing Constant, μ = 0.4

Seasonal Effects In Stock Markets

Overtime, seasonal and calendar-related trends have developed that are well known and anticipated by most investors. Even though the effects of these economic conditions are debatable, understanding seasonal effects is important because they often explain changes in the markets that are not attributable to prevailing economic and business circumstances. During the Christmas period, three phenomena are closely related which are;

The December Effect: This phenomenon occurs during the weeks immediately before and after Christmas in which markets will decline.

The reason is that many traders and investors will **sell off shares just before the end of the year**, in order to claim a capital loss for the year and reduce their taxes.

January Effect: This is a market rally in the first week of January, **when shares sold by traders at the end of the year are repurchased.** These effects in average monthly returns for stocks were half a percent higher in January than in the other 11 months.

Santa Claus recovery: This occurs in years where, the December effect is strong, early in the month of December. This effects in a bit of a recovery in average returns for stocks, due to share buying by bargain hunters.

Implications of Seasonality At End of The Year

- The December effect will cause market values to decline because investors will sell their losing positions at the end of December to obtain tax losses.
- Investors predicate that these stocks sold off will be at a discount to their market value in the first few weeks of January.
- The January effect will bring in recovery in the market due to bargain hunters stepping in and buying the sold stocks, hence creating market pressure on the investors.
- Due to the December effect, small cap[4] stocks will likely outperform large caps[5] during the middle of December and **many large cap stocks will do poorly in this season**
- **Conclusion:** It is advisable to invest in **small cap stocks** during this period of the year.

Popularity of the Company

The popularity of a stock can affect its market price. However, popularity is not necessarily a predictor that the stock price will increase or decrease. In order to statistically consider popularity, the trading volumes of the stocks were considered. Trading volume is the number of shares bought and sold between investors over a particular period.

High trading volumes indicate investor's interest in a stock. Coupled with an increase in demand and supply, these stocks may have an increase in liquidity[1] and narrower bid[2] spreads. The spread is the difference between the ask[3] and bid price of a stock. These are positive stock attributes and contributed to the selection of each stock for our investment.

The Trading volumes and stocks volatility gives us better forecast views of the market. Stocks that trade in high volumes tend to be less volatile than those trading at lower volumes. Stocks with higher trading volumes are less volatile because investors can buy, sell and trade large amounts of the stock without the transactions having a significant effect on the stock price. For long-term investors, stocks that are less volatile are more attractive than those that are more volatile. This decreased volatility tends to increase the price of the stock. Stocks with lower trading volumes however, are more volatile.

Company Profiles

Company Name	Ticker	IPO Year IPO Price	Initial Year 02/02/2015
Celgene Corp	CELG	**03/26/1990** $8.25	$118.57
MediciNova, Inc	MNOV	**12/08/2006** $12.10	$4.22
Clorox Company	CLX	**03/21/1983** $29.25	$107.83
Tesaro, Inc	TSRO	**06/28/2012** $13.69	$39.76
Biogen Idec Inc	BIIB	**09/17/1991** $18.75	$390.0

Celgene Corp (Nasdaq)

- Pharmaceutical company that develops and sells drugs to treat blood cancer with main drugs Thalamid and Revilmid.
- Company is generating high levels of growth with total product sales yearly rise of 19% and topping $7 billion in yearly revenue 2014
- High trading volume of 4.23 million and market capitalization of $95 billion

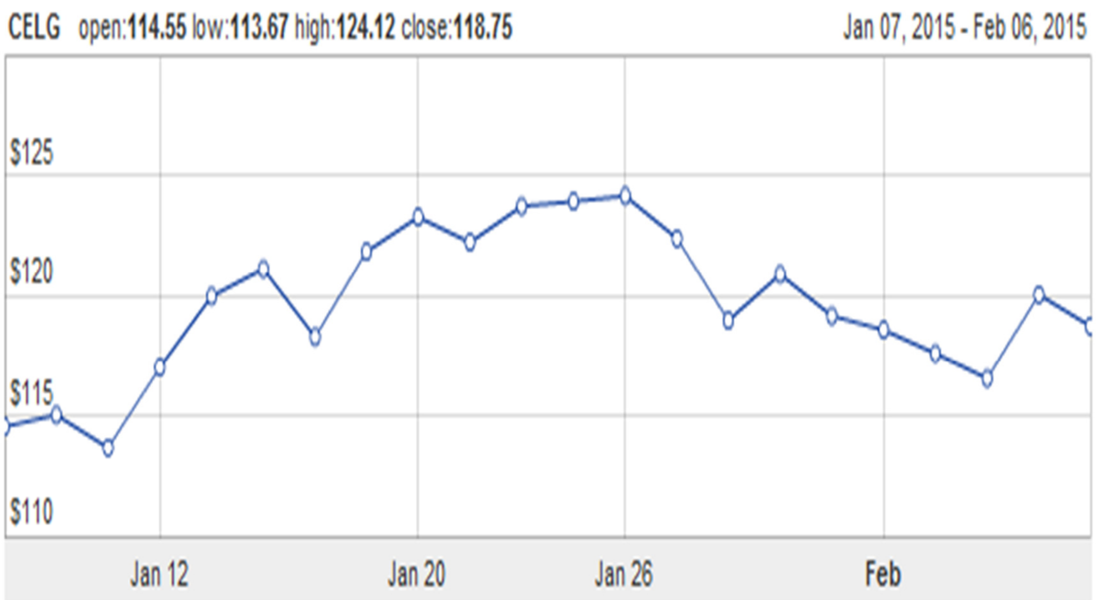


Figure 8a

MediciNova, Inc (Nasdaq)

- Fast growing US Biopharmaceutical Company that specializes in novel therapeutics.
- On February 3rd, it received a Notice allowance from the US Patent and Trademark Office for a Pending patent application for
- Patent will expire no earlier than December 2032 and the company expects to boost

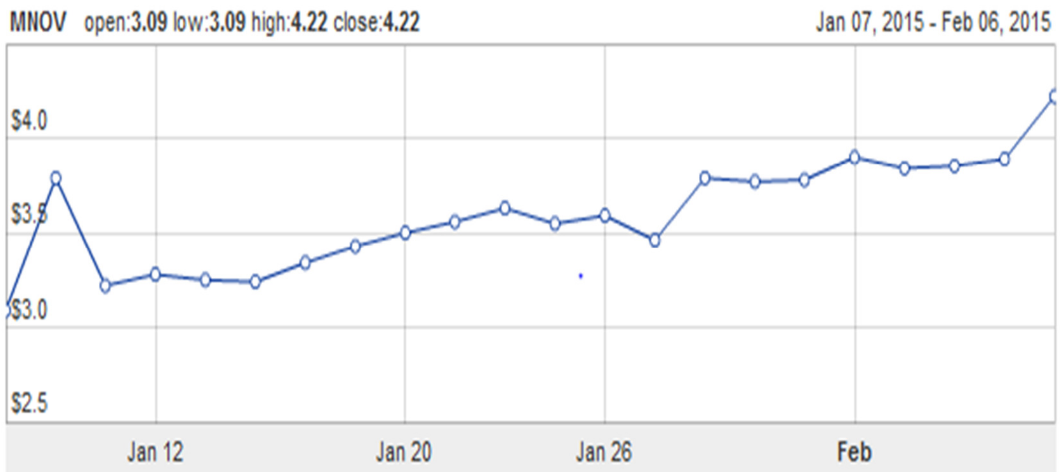


Figure 8b

Clorox Company (NYSE) :

- US Multinational Manufacturer of cleaning products and water filters.
- Company has posted strong quarterly results since January 2nd.
- While Clorox is working to expand its presence to Latin American countries, the economic crisis in Venezuela will make it even more influential

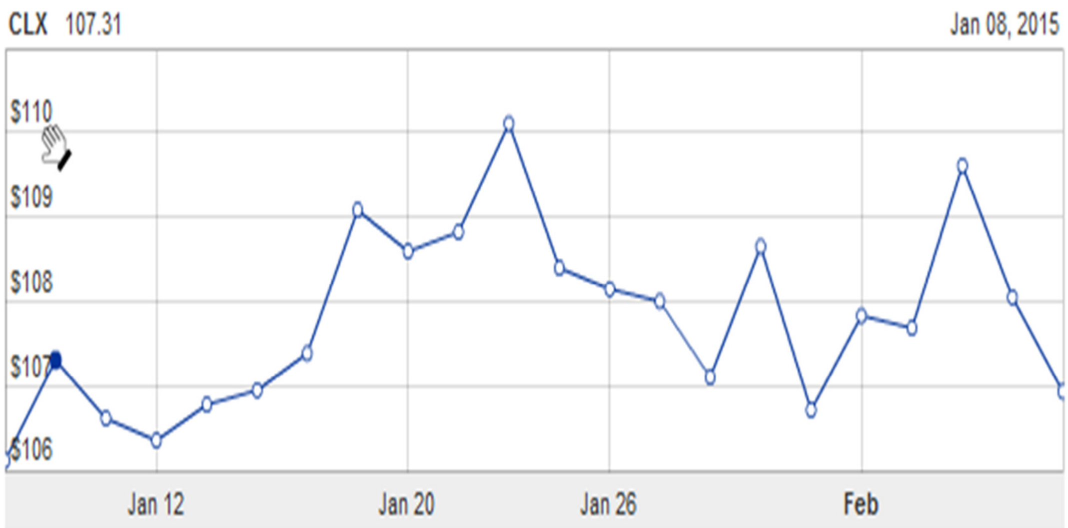


Figure 8c

Tesaro, Inc (Nasdaq) :

- TESARO is a small privately held oncology-focused biopharmaceutical company based in Boston, MA
- Company to Announce Fourth-Quarter 2014 Financial Results on February 19, 2015
- Large market capitalization of $1.35 billion

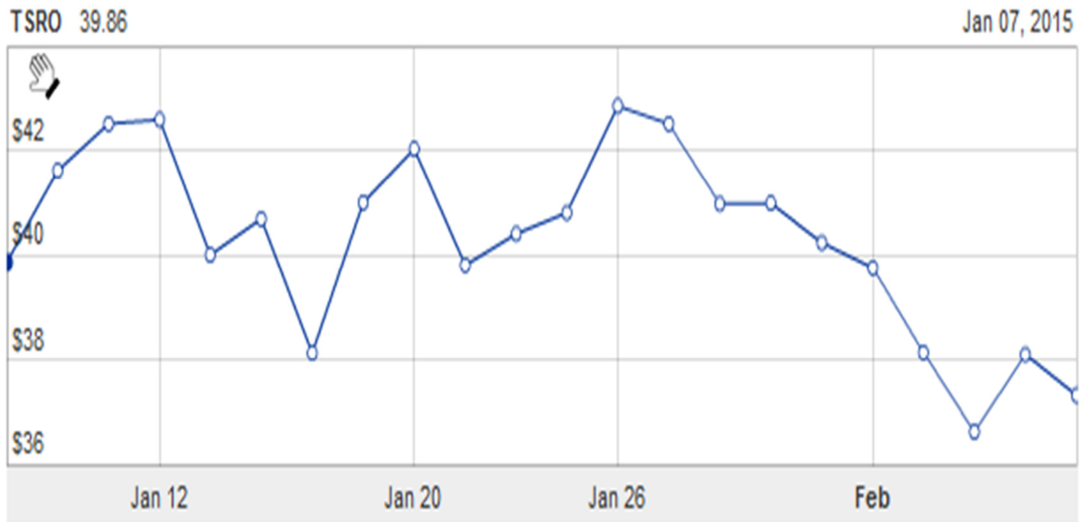


Figure 8d

Biogen Idec Inc (Nasdaq)

- Company produces drugs for cancer and neurological treatment. Main drugs are Avonex and Tecfidera.
- Company boasted a 16% revenue growth in January quarter largely due to outstanding sales of its twice-daily MS pill Tecidera
- Tecidera may secure a company revenue of about $ 4billion by December 2015

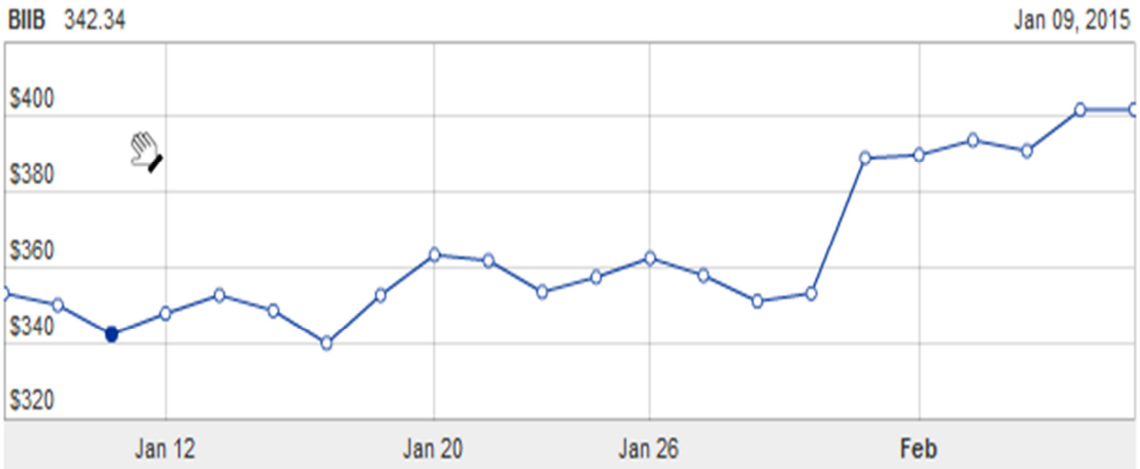


Figure 8e

Results and Analysis

The MATLAb ARIMA steps used to generate each ARIMA prediction is shown below

```
% ARIMA Model

 Mdl=arima(2,1,2);  % Arima 2,1,2 model

 EstMdl=estimate(Mdl,x); %Estimate Arima Parameters from input data

[arimaPredictions arimaErr]=forecast(EstMdl,20,'Y0',x); %Forecast 20 future prices

 [Resids Varies] = infer(EstMdl,x); %Infer past results (estimate model of past data)

arimaplot = [(Resids+x)' arimaPredictions']; %vector of past and future results to plot

(this is the completed model)
```

Week 1

Stock Name	Symbol	Price on Feb 2nd	Price on Feb 6th	Volume Bought	Amount Feb 2nd	Amount Feb 6th
Celgene Corp	CELG	118.57	118.75	100	11857	11875
Tesaro, Inc	TSRO	39.76	37.32	450	17892	16794
MediciNova Inc	MNOV	4.22	3.90	700	2954	2730
Clorox Company	CLX	107.83	106.93	400	43132	42772
Biogen Idec Inc	BIIB	390.0	402.00	60	23400	24120

Table 5: Stock Performance For First Week

- Gross Budget = $ 100,000
- Amount Invested On 02/02/2015 = $99,235
- Total Price Value of Stocks on 02/06/2015 = $98,291
- A loss of **$944** was made on 02/06/2015

Week 2

Stock Name	SYM	Price Feb 6th	Price Feb13th	VOL	Amount Feb 6th	Amount Feb 13th
Celgene Corp	CELG	118.75	115.88	100	11875	11588
Tesaro, Inc	TSRO	37.32	38.49	450	16794	17321
MediciNova Inc	MNOV	3.90	3.62	700	2730	2534
Clorox Company	CLX	106.93	109	400	42772	43600
Biogen Idec Inc	BIIB	402.00	391.66	60	24120	23499

Table 6: Stock Performance For Second Week

- Gross Budget = $ 100,000
- Amount Invested On 02/02/2015 = $99,235
- Total Price Value of Stocks on 02/06/2015 = $98,291
- Total Price Value of Stocks on 02/13/2015 = $98542
- We registered a loss of **$693** in February 13th. It is an improvement from the $944 registered on February 6th

Week 3

Stock Name	SYM	Price on Feb 2nd	Price Feb 13th	Price Feb 20th	VOL	Amount Feb 13th	Amount Feb 20th
Celgene Corp	CELG	118.57	115.88	123.60	100	11588	12360
Tesaro, Inc	TSRO	39.76	38.49	44.20	450	17321	19890
Medici Nova Inc	MNO V	4.22	3.62	3.51	700	2534	2457
Clorox Compa ny	CLX	107.83	109	108.94	400	43600	43576
Biogen Idec Inc	BIIB	390.0	391.66	408.05	60	23499	24483

Table 7: Stock Performance For Third Week

- Gross Budget = $ 100,000
- Amount Invested On 02/02/2015 = $99,235
- Total Price Value of Stocks on 02/06/2015 = $98,291
- Total Price Value of Stocks on 02/13/2015 = $98,542
- Total Price Value of Stocks on 02/20/2015 = $102,766 We registered a profit of $2766 on February 20th. This is an improvement from the $1458 loss we registered on February 13th

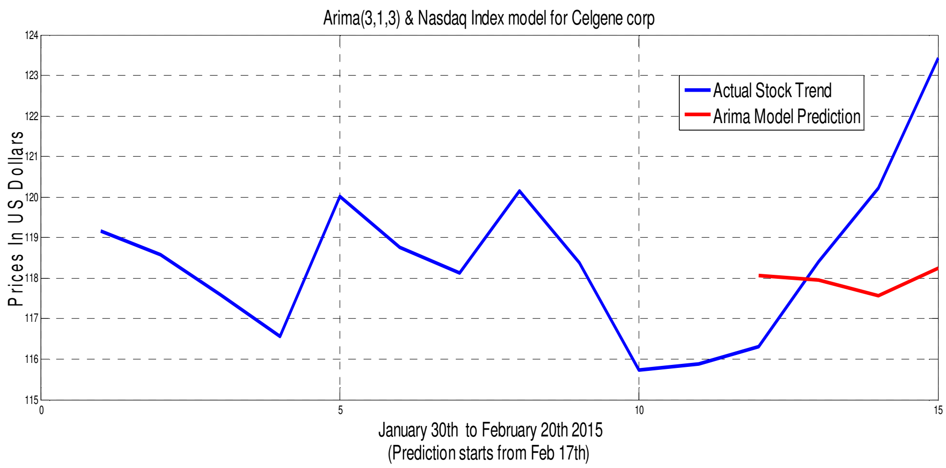


Figure 9a:Stock Prediction for CELG

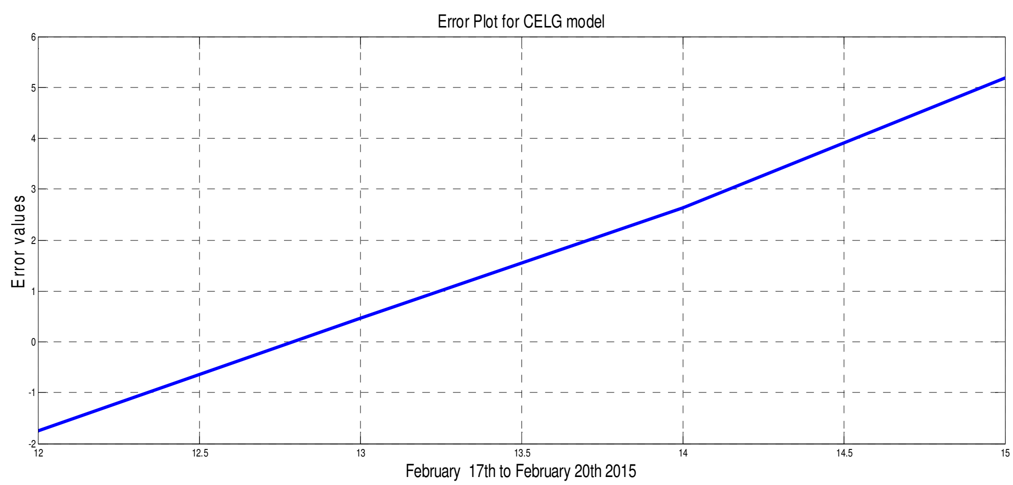


Figure 9b : Error plot for CELG model

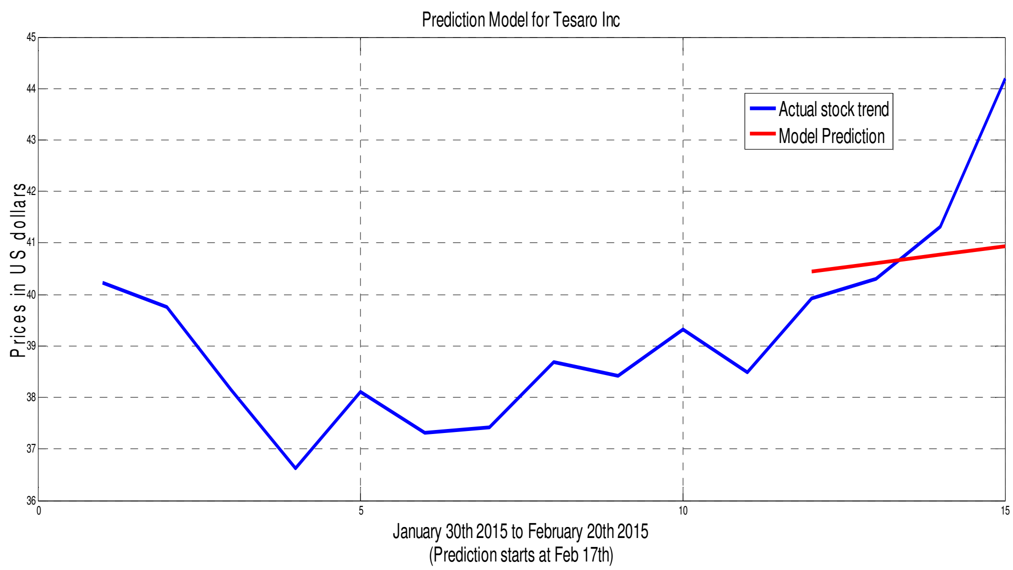


Figure 10a : Stock Prediction for Tesaro Company

Figure 10b : Error plot for Tesaro model

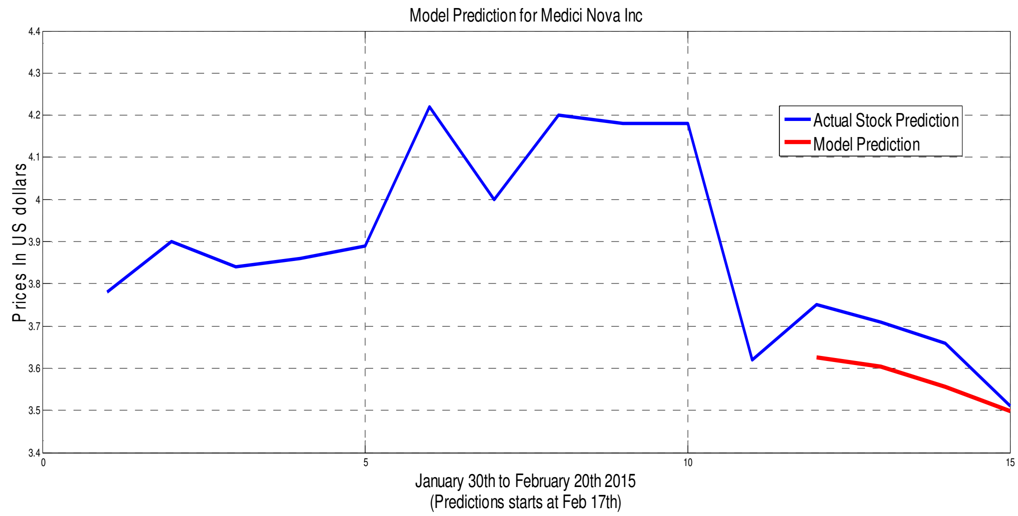


Figure 11a : Stock Prediction for Medici Nova Company

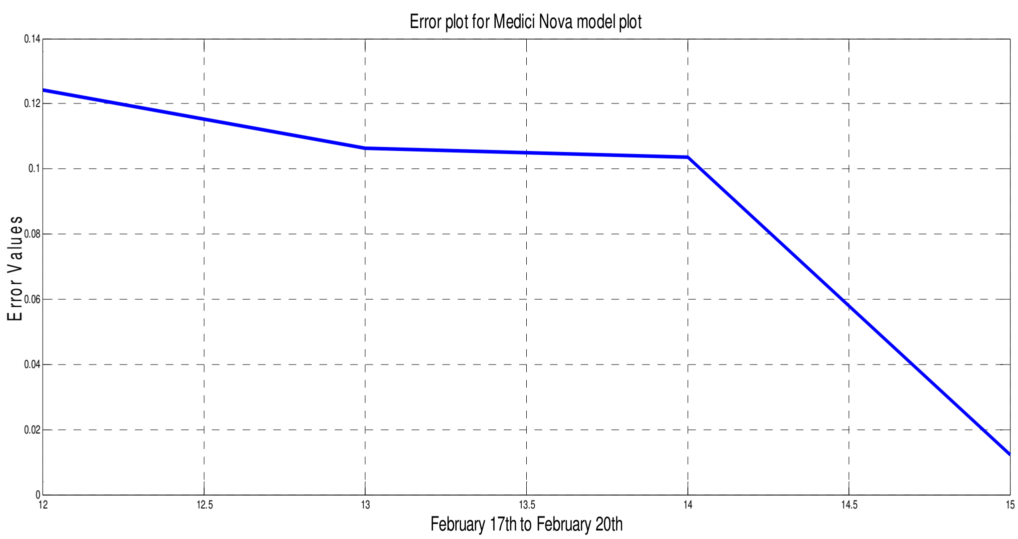


Figure 11b : Error plot for Medici Nova model

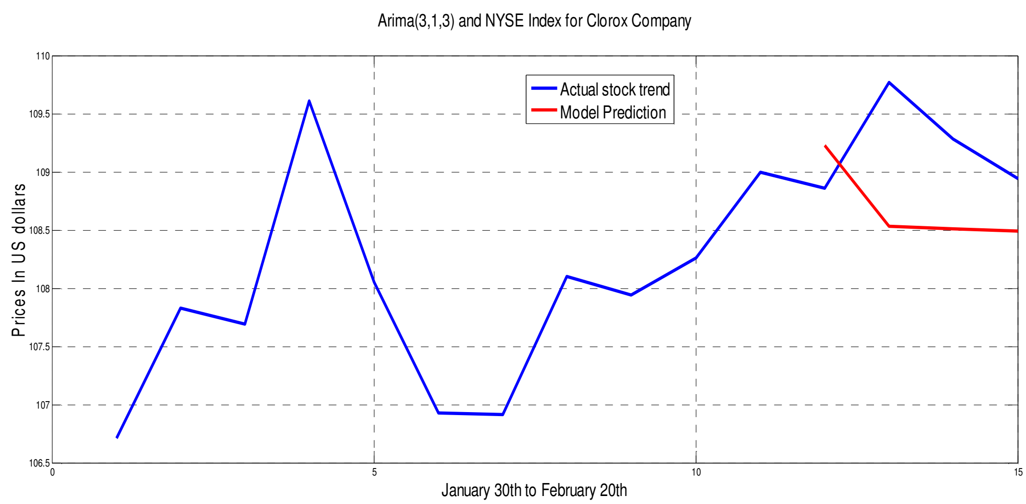


Figure 12a : Stock Prediction for Clorox Company

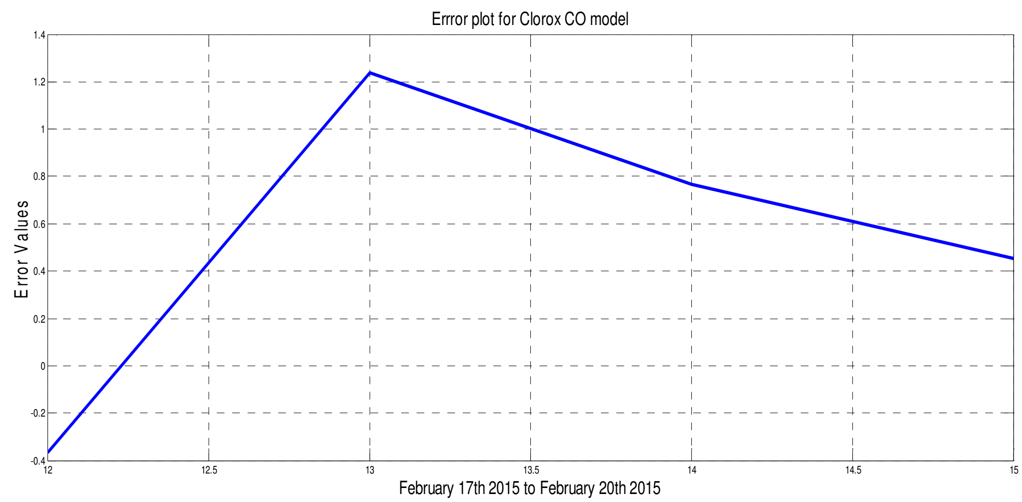


Figure 9b : Error plot for Clorox model

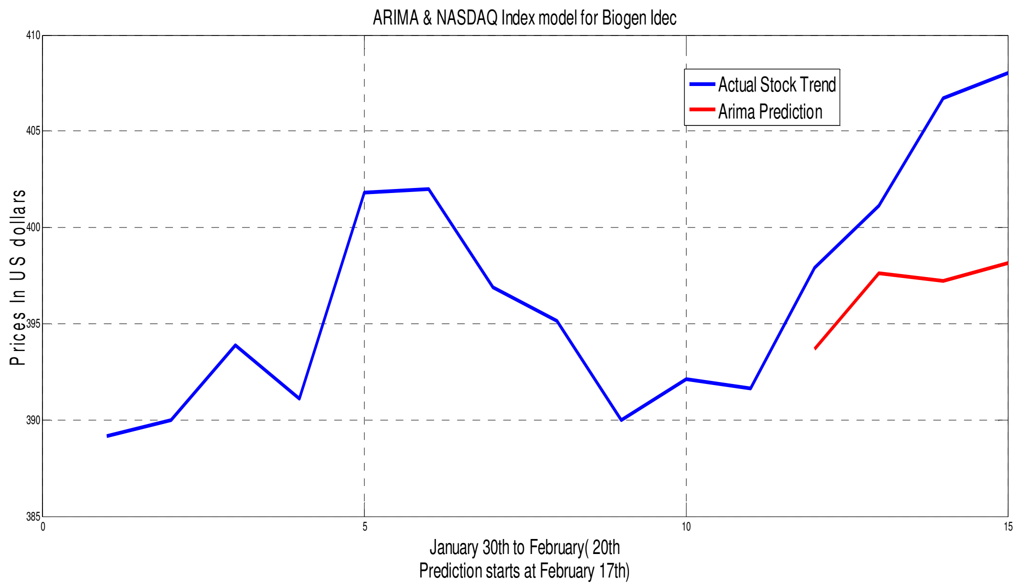


Figure 13a : Stock Prediction for Biogen Idec Company

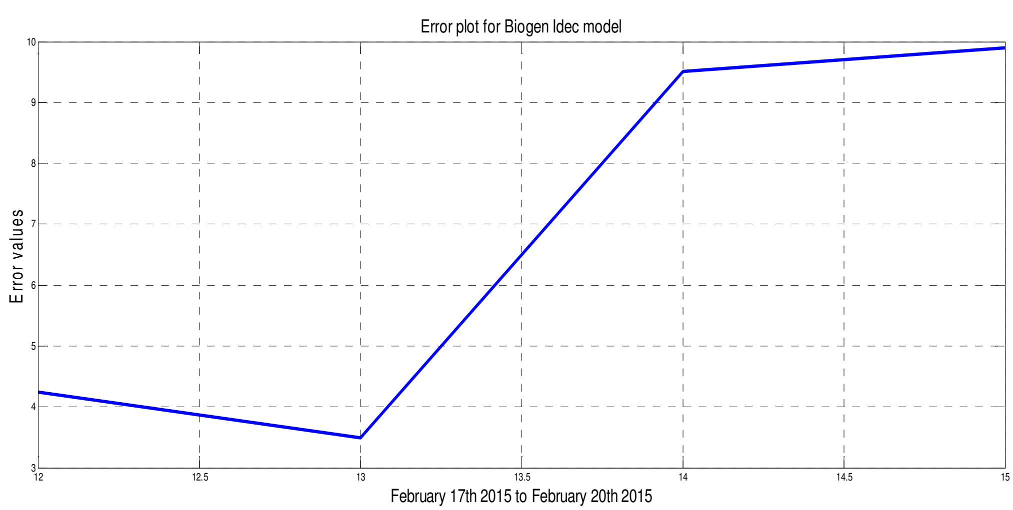


Figure 13b : Error Plot for Biogen Idec Company

Problem and Solution

Through observing each of the stock prediction graphs above, we can see that the stock prediction model worked fairly well in predicting the trends.

Celgene Company

Prediction Analysis: The stock prediction plot for Celgene predicted a sharp price increase from February 17th to February 20th. Nevertheless, this sharp price increase is at a lower rate than the actual stock price. This is not a negative observation since we will actually make a greater profit than expected.

Tesaro Company

Prediction Analysis: The prediction plot was a linear line which closely followed the actual trend during the first two day, then deviated from it in the next days. We can observe that the model predicts a price increase in the stock value and the actual stock price is even of a greater value. This Is a desirable result because we will register more profit than forecasted by the model.

Medici Nova

Prediction Analysis : The prediction curve accurately predicted a fall in the stock prices as shown in the figure.

Clorox Company

Prediction Analysis : The fitting curve did not correctly depict the future trend. In the first two days, the fitting curve predicted a sharp fall in prices and later showed a moderate fall in the prices of the stocks. This is different from the actual trend that showed a sharp increase in the price of the stock, then a sharp decrease after two days. Nevertheless, the actual stock price is greater than that predicted, so we will still had a stock profit.

Biogen Idec Company

Prediction Analysis: The actual trend and the prediction are fairly accurate as they both portray a price

Increase in the stock price. The actual trend increases at a greater rate than the prediction. We can say that our stock model worked well with Biogen stocks.

Summary

We invested $100,000 in the purchase of five biotechnology stocks. Our stock prediction model was then used to forecast the stocks trend in a monthly period, aiming to generate profits.After the first week of prediction on February 6th 2015, we registered a price loss of $98,291. The next week, on February 13th, we registered a smaller price loss of $98,542. On the third week, we had made a profit of $102,766.

Conclusion

The goal of this IQP project was to develop a simple, efficient and inexpensive forecasting tool that will assist new and inexperienced investors in their stock market businesses. We developed our stock model, by first building up our regression model, then adding this model to the ARIMA formula. In order to introduce the effect of seasonality, our stock model was fractionally added to a similar forecast for the market exchange index. This gave us our final stock prediction model. This final model was then used to invest $100,000 into 5 biotechnology stocks and forecast stock profits in a monthly period. By the end of our monthly period, a stock profit of $102,766. The stock model followed trends close to the actual stock trends even though, the actual stocks always moved at greater rates. We can confidently conclude that our model was able to work for our $100,000 investment.

Even though our model worked in generating an investment profit, we cannot exclude the many unpredictable events that affected our results. Three of the biotechnology companies had received patents for the early developments of cancer drugs. This may have increased their stocks hype, hence boosting investor's interest and the stock prices. The falling oil prices may have also positively affected these biotechnology companies because low oil prices will lower the companies cost of production and enable investment in other sections of the company. We see that such factors were not included into the model. Other stock stochastic factors like market capitalization and stock traded volumes could not be effectively quantized in our final Stock model due to the short length of our research period.

Our advice for future stock investors is to strive in mathematically including stock parameters like, stock price, stock traded volume, market capitalization and even investors rating into their stock models. Our model included only the stock prices and produced positive results for the prediction of 5 biotechnology stocks. We will not recommend our model for the forecasting of other types of stocks because biotechnology companies behave in differently from other stocks.

Finally, since the price of the stock is strictly tied to thousands of unpredictable events, it is recommended to proceed with caution when using any stock prediction tool and with this prediction tool in hands, getting educated with other prediction tools will be a big help. This allows a comparison of tool to use and avoid uninformed risks taking.

Acknowledgment

I would like to thank Professor Humi Mayer for his constant encouragements and quality advising during the course of this research. Through his guidance, I bolstered my mathematical knowledge by learning the behavior of hermit polynomials and the nature of ARIMA models. I also sharpened my MATLAB skills and I developed a great interest in the stock market. I appreciate the critics he made about my presentation styles because they helped me to improve them and made me capable of facing future reaseach challenges. I am thankful and grateful for all his dedication to see me make it to the end of this project. Thank you, Professor Humi Mayer.

References

OLSEN, R. A. (1998). Behavioral finance and its implications for stock-price volatility. *Financial Analysts Journal, March/April,* 10-18

[1] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994.

Zhouxiao WU, Jialia Gao : How to predict Price of Stock.

FAMA, E. F., Fisher, K. L., Jensen, M. C., & Roll, R. (1969). The Adjustment of Stock Prices to New Information. *International Economic Review, 10*

G-S. Jang *et al.* Intelligent stock trading system with price trend prediction and reversal recognition using dual-module neural networks J. Applied Intelligence, (1993).